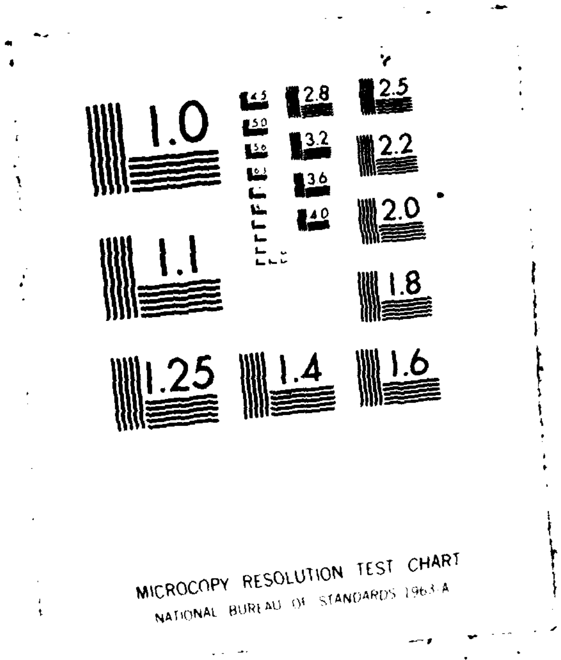
AD-A179 049　A MODIFIED GOODNESS-OF-FIT TEST FOR THE LOGNORMAL DISTRIBUTION WITH UNKNO..(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGI.. L I WHITSEL　1/2

UNCLASSIFIED　DEC 86 AFIT/GOR/MA/86D-6　F/G 12/1　NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

AFIT/GOR/MA/86D-6

A MODIFIED GOODNESS-OF-FIT TEST FOR THE LOGNORMAL DISTRIBUTION WITH UNKNOWN SCALE AND LOCATION PARAMETERS

THESIS

Lynnette Townsend Whitsel
Captain, USAF

AFIT/GOR/MA/86D-6

AFIT/GOR/MA/86D-6

A MODIFIED GOODNESS-OF-FIT TEST FOR THE LOGNORMAL DISTRIBUTION WITH UNKNOWN SCALE AND LOCATION PARAMETERS

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
in Partial Fulfillment of the
Reguirements for the Degree of
Master of Science in Operations Research

Lynnette Townsend Whitsel, B.S.
Captain, USAF

December 1986

PREFACE

This thesis develops modified critical value tables for the lognormal distribution using the Kolmogorov-Smirnov, Anderson-Darling, and the Cramer-Von Mises goodness-of-fit tests. These critcal value tables can be used when the scale and location parameters are estimated from the observed data. Next it compares the power of these new tests when the hypothesis being tested is the lognormal. The data tested come from the lognormal, Weibull, gamma, beta, exponential, and normal distributions. Finally this research determines the relationship between the modified critical test statistics and the known shape parameter.

There are several people I would like to thank for their various contributions to this thesis effort. First, for suggesting the topic and for being my advisor, I wish to thank Dr. Albert Moore. I also wish to thank Dr. Joseph Cain for being my reader.

Finally, and most of all, I thank my husband, Kent, for staying with me during our first year of marriage while at AFIT. Without his understanding and patience, this thesis would not have been possible.

Lynnette Townsend Whitsel

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract

This thesis developed modified goodness-of-fit tests for the three parameter lognormal distribution when the location and scale parameters must be estimated from the sample. The critical values were generated for the Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Cramer-von Mises (C-VM) goodness-of-fit tests, using the Monte Carlo methods of 5000 repetitions, to simulate samples of size 5,10,...,30 and the shape parameter ranged from 1.0 to 4.0 in increments of .5.

The second part of the research also involved a Monte Carlo simulation of 5000 repetitions for sample sizes of 5, 15, and 25. From these observations, the power of the test was determined by counting the number of times the modified goodness-of-fit tests incorrectly accepted null hypothesis that the distribution was lognormally distributed. The data used in this power comparison came from the lognormal distribution (shape = 1.0 and 3.0), Weibull, gamma, beta, exponential, and normal distributions.

The third and final phase of research was to determine the functional relationship, if any, between the known shape parameter and the new modified critical values. This was completed by using SAS.

A MODIFIED GOODNESS-OF-FIT TEST FOR THE LOGNORMAL DISTRIBUTION WITH UNKNOWN SCALE AND LOCATION PARAMETERS

I. INTRODUCTION

The Air Force and other branches of the military are placing an increased emphasis on system reliablity and maintainability. In studying current systems, statistics are gathered and used to determine mean-time-to-failure (MTTF), mean-time-to-repair (MTTR), and expected life of the weapon. Statistics are also used in the research of proposed systems, by predicting MTTF and MTTR of the new parts and thus, predict the reliability of those parts. The statistics gathered are then classified as particular distributions.

The Air Force uses these various distributions in their simulation models to predict and study factors and effects. These studies fit data such as time-to-failure of equipment components, maintenance service times, nuclear fallout particles, and error clusters in communication circuits (28:3-4). With the current budget constraints, all branches of the military are interested in cost effectiveness of new systems. Aitchison's book on the lognormal distribution, printed and used by Cambridge University, highlighted the distributions numerous applications to economic problems.

BACKGROUND

A distribution is a single or multi-parameter theoretical, statistical model of data, often used to predict the behavior of a population of entities by studying a sample (portion) of the population. Goodness-of-fit tests measure the correlation between this observed data sample and a particular statistical distribution. The four most often used goodness-of-fit tests are the Chi-square, the Kolmogorov-Smirnov (K-S), the Cramer-von Mises (C-VM), and the Anderson-Darling (A-D).

Before applying any goodness-of-fit test, the researcher must complete four steps to determine which distribution is suggested by the data. First, the analyst must collect data for the study problem. Second, he must hypothesize (guess) which statistical distribution best characterizes the data. Next, he must estimate parameters as suggested by the data. The analyst then uses one of the above goodness-of-fit tests to determine if the data follows the statistical distribution as hypothesized. If the test rejects the hypothesis, he must then return to the second step and try another distribution.

This study of statistical data is becoming more frequently used in the Air Force for various problems. For example, in light of budget cuts, the Air Force is more concerned than ever before with studying reliability and maintainability (R&M) of systems and all parts of these systems. In studying R&M, analysts use observations in a graphical form to estimate statistical distributions suggested by the sample data. From this information the analyst can determine such statistics as mean time between failure (MTBF) and mean time to repair (MTTR). The difficulty with using this graphical representation of data is that the estimates of the parameters are less accurate.

PROBLEM STATEMENT

Currently, there is no test to determine goodness-of-fit for the lognormal distribution when the scale and location parameters are unknown. When a random sample of data is collected, a test could be used to determine if the population of data was taken from this type of distribution. The problem to be solved is to apply current goodness-of-fit tests to lognormal distributions with no known scale and location parameters.

RESEARCH QUESTION

This research is to develop a modified goodness-of-fit test for the lognormal distribution with unknown scale and location parameters.

OBJECTIVES

To solve this problem, the analyst must accomplish three objectives. First, critical value tables for the lognormal distribution must be generated and documented for each of the two modified goodness-of-fit tests. These modified tables are used when the scale and location parameters are unknown. Next, the powers of the tests are compared, to determine the best test to use for the lognormal distribution with unknown parameters. This power is the probability that the statistical test will correctly reject a wrong distribution guess. Last, the analyst must determine functional relationships, if any exist, between the shape parameter and the goodness-of-fit statistics. This relationship allows one to interpolate missing values not documented in the generated tables.

II. GOODNESS-OF-FIT TEST

Introduction

Goodness-of-fit tests measure the correlation (agreement) between an observed data sample and a particular statistical distribution. Normally, a goodness-of-fit test is used to examine a random sample to determine if the data is from a hypothesized specific function (28:2-1). If, given a certain level of confidence, the test indicates a close fit, the sample data is from a specified distribution, this distribution can be used in simulation modeling to represent real world occurrences. The Air Force is using simulation models more and more to help managers answer "what-if" type questions. Other uses for these simulation models are: how to determine which systems to buy and projection of maintenance figures for future systems based on sample data from prototypes.

Background

One branch of statistics is devoted to the study of distributions that do not depend on certain parameters being known. This branch is known as non-parametric statistics and the tests statistics developed for these studies are non-parametric or distribution-free tests (23:68).

The Chi-square test for goodness of fit, first presented in 1900 by Pearson, is the oldest and most well known goodness-of-fit test (7:189). To use the Chi-square test, one compares the frequency of the observed data with the expected frequencies of the hypothesized distribution (28:2-2). The Chi-square is the most flexible test when dealing with unknown parameters; for each parameter unspecified, a deqree of freedom is subtracted. With this flexibility, comes certain drawbacks; as more parameters are estimated, the power of the test is diminished greatly. The lower the power, the greater the possibility that the test will accept a false hypothesized distribution with greater frequency. The second drawback of using the Chi-square test is it's use is only valid for large sample sizes; more than 50 (28:272) and it requires the data to be arbitrarily grouped (28:2-2) which may affect the results.

Another often used goodness-of-fit test for distribution-free test is the Kolmogorov-Smirnov (K-S) test, introduced by Kolmogorov in 1933 (7:344). Kolmogorov and Smironov developed their goodness-of-fit test to use the maximum distance between the observed data and the hypothesized distribution to measure how close the functions resemble each other (7:344). The K-S test statistic enables one to form "confidence bands" for different levels of confidence, about the hypothesized distribution (7:346). If the data lies within the bands, the data is accepted as fitting the hypothesized distribution. The drawback with using the standard K-S test is that all parameters must be specified; there can be no unknown parameters that must be estimated from the sample (28:2-2).

A third goodness-of-fit test that measures distance between the hypothesized CDF and the observed data is the

Cramer-von Mises test. This test is based on the squared integral of the distance between the observed data (in the form of an empirical distribution function which is discussed later) and the distribution to be tested (28:2-12).

A member of the Cramer-von Mises family of goodness-of-fit tests is the Anderson-Darling test statistic. Anderson and Darling wanted more flexability in testing goodness-of-fit, thus they introduced a technique of incorporating a weight function into the K-S and C-VM test statistics (28:2-13). This weight function counteracts the decreasing difference between observed data and hypothesized distribution, at the tails. In effect, it heavily weights the difference at the tails.

In 1948, David and Johnson (8) furthered the study of non-parametric statistics when they discovered that a distribution having only a location and scale parameter, can have these parameters replaced with invariant estimators, without affecting the goodness-of-fit test results. These estimators are invariant in that if x is transformed by $x=ax+b$ the estimate $T=T(x)$ is also transformed (i.e. $T=aT+b$) (28:2-3). From these results, it has been found that critical values based only on sample size and significance level can be generated (39:5). This principle can be extended to three-parameter distributions given that the shape parameter is held constant. In these modified tests, the test statistic is unchanged but estimates are used in the place of known parameters.

Hypothesis Testing and Test Statistics

Before studying statistical distribution and goodness of fit, one must have a working understanding of the basic concept of hypothesis testing. The first step of this testing procedure is to observe and gather data on a portion (sample) of the population. From this sample, the analyst attempts to draw conclusions on the behavior of the parent population. The next step is to hypothesize (guess) what theoretical distribution best fits the observed data. The analyst then chooses a test to determine if the data does indeed come from the theoretical (null hypothesis) distribution. Using the critical value formula for the test chosen, the analyst follows the test to either accept or reject whether the data fits the hypothesized distribution.

There are two possible results of hypothesis testing: to accept a stated distribution guess (null hypothesis) or to reject this distribution. From these two outcomes, there are two types of errors that can be made. The Type I error, denoted α (alpha), is the probability of rejecting the null when it is correct and the Type II error, denoted β (beta), is the probability of accepting when the null is incorrect (28). Accepting the null hypothesize (denoted Ho) does not prove that it is true; there was simply insufficient evidence to reject the alternative hypothesize. Accepting the alternative hypothesis is the same as rejecting the null; that is, there is significant evidence that Ho is false.

The strength of a goodness of fit test is measured by the power of the test. The greater the probability of rejecting a false null hypothesis (denoted by $1-\beta$), the more powerful the test. (7:79)

Empirical Distribution Function (EDF)

Since the true distribution of observed data is almost never known, one must often make an educated guess about the parent population from the sample statistics, based on the empirical distribution function (EDF). The EDF is often used to compare this observed data to a hypothesized distribution function (28:2-6). From this "sample" graph, estimates can be made about the unknown distribution of the population H(x) by using the EDF.

The empirical distribution function S(x) is the function of X equal to the fraction of X's that are less than or equal to X for each X between negative infinity and positive infinity for the random sample: X1 ,X2 , ... Xn. That is:

$$Sn(X) = \frac{\text{number of values} < x}{\text{total number in sample}} \qquad (1)$$

For a sample of n size, the EDF is denoted as Sn(x).

The EDF is always a step function, with each step height equaling 1/n, with the EDF a non-decreasing function from zero to one (28:2-7)

$$Sn(X) = \begin{cases} 0 & \text{for all } X < X_1 \\ i/n & \text{for } X_i < X_{i+1}, \quad i=1,2,\ldots,n-1 \\ 1 & \text{for all } X > X_n \end{cases} \qquad (2)$$

The Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and the Cramer-von Mises (C-VM) tests for goodness of fit are of the EDF type (34:730).

Modified Goodness-of-Fit Tests

The current goodness-of-fit test can be used with the various distributions, "provided there are no unknown parameters in the hypothesized distribution." (39:1) Currently there are modified goodness-of-fit tests for the Weibull, (4;22) Normal, (20;31) Gamma, (34:37) Pareto, (28) Logistic, (38) Exponential, (21) and the Uniform (35) distributions.

Woodruff used a Monte Carlo technique to generate sample observations as well as to develop critical value tables for the K-S, A-D, and C-VM tests for gamma distribution with unknown parameters. Similar methods were used in the other papers in order to develop the modified goodness-of-fit tests.

Chi-Square Goodness-of-Fit Test

The Chi-square test can be used for large sample sizes (ie. more than 50) where the distribution is either discrete or continuous. This test is particularly well suited for use when parameters are estimated from the sample, by maximum likelihood techniques (33:730). The Chi-square procedure to determine goodness-of-fit begins with placing the observations in intervals, or "cells". The test statistic is as follows:

$$\chi_0^2 = \sum_{i=1}^{k} (Oi-Ei)^2/Ei \qquad (3)$$

"where Oi is the observed frequency in the ith class interval and Ei is the expected frequency in that class interval." (3:350) The null hypothesis (Ho) for this test is that the data conforms to the assumed distribution. The critical value for the test is found in Chi-square tables. If the test statistic is greater than the value found through calculation, reject Ho.

Kolmogorov-Smirnov

In Massey's article on the K-S test, he describes the procedure involved in determining goodness-of-fit with this test is to "draw the hypothetical cumulative distribution function on a graph and to draw curves a distance above and below the hypothetical curve." (23:69) If the observed distribution, Sn(X), passes outside of this drawn curve at any point, this distribution is rejected as not fitting the data (3:269-271).

There are several advantages to using this test to fit data to a theoretical distribution. First of all, the KS test is particularly useful when the sample size is small. It also appears to be more powerful than the Chi-square test for sample data of any number; however, the standard K-S test is only applicable when parameters are known about the data (20:399). The K-S Test statistic is the largest (denoted "sup" for supremum) vertical distance between the hypothesized distribution F(X) and the observed EDF, Sn(X) (28:2-11). The test statistic is

$$D = \sup_{x} \left| F(X) - Sn(X) \right| \tag{4}$$

The equivalent computational form is

$$D = \max \left(D+,\ D- \right) \tag{5}$$

where Ho is rejected at the given level of significance, if D is greater than the critical value given at that level (7:358).

<u>Anderson-Darling</u>

The Anderson-Darling test statistic is a subset of the C-VM family of statistics. The unique feature of this test is the incorporation of a weighting function into the K-S and C-VM test statistics (28:2-13). The A-D test statistic (2:767) is as follows:

$$A^2 = n \int_{-\infty}^{\infty} [Sn(X) - F(X)]^2 \, \theta [F(X)] \, dF(X) \tag{6}$$

where $\theta [F(X)] = F(X)*(1-F(X))^{-1}$. Its computational form is

$$A^2 = -n - (1/n) \sum_{i=1}^{n} (2*j-1) [\ln Z_j + \ln(1-Z_{n+1-j})] \tag{7}$$

where X1 < X2 < ... < Xn are n ordered observations from a sample and Zj = F(Xj) for j = 1, 2,...n (28:2-14)

Cramer-von Mises Test

The Cramer-von Mises test is based on the squared integral of the difference between the observed data, Sn(X), and the distribution being tested, F(X), (28:2-12). The C-VM test (2:766) statistic is derived by the following formula:

$$w^2 = n \int_{-\infty}^{\infty} [Sn(X) - F(X)]^2 \, dF(X) \qquad (8)$$

and also its computational form:

$$w^2 = [1/(12n)] + \sum_{i=1}^{n} [Z - (2*i-1)/2n]^2 \qquad (9)$$

where X1 < X2 <...< Xn are n ordered observations from the sample. The C-VM goodness-of-fit can be considered a special case of the A-D statistic where $\theta[F(X)] = 1$ (28:2-13).

III. Lognormal Distribution

Introduction

"The lognormal distribution in its simplest form may be described as a distribution of a variate whose logarithm obeys the normal law of probability." (1:1) Although the lognormal distribution has not been studied as long as the normal distribution, it's origin can be traced as far back as 1879 (1:1). The lognormal, by its very nature, has many properties which are derived from the normal distribution. There are also those properties possessed by the lognormal which cannot be easily, if at all, found in normal theory.

History

Probably the most used distributions in statistics is the normal distribution curve, developed by Gauss in 1809 (18:6). This curve could almost but not completely describe certain distributions observed by statisticians of the day. During the late 1800s, attempts were being made to discover and construct systems of frequency curves that represented a wider variety of distributions than the normal distribution curve could describe. These new systems varied from normal in their skewness; thus they were refered to as "skew frequency curves" (16:149).

K. Pearson, in 1985, and Charlier, in 1905, appear to have completed two of the more successful attempts at constructing these skewed systems. In 1898, Edgeworth proposed the concept of transformation which he termed "method of translation". Since most work at the time dealt with the normal distribution, the natural course was to relate the new system to current work on the normal. This method sought a function of an observed random variable which was closely related to the normal random variable. Normal theory was then used on the new "transformed" variables (18:6). Edgeworth's method was not generally accepted due to the lack of variety of shapes it could be used for. This technique did however help to further the studies of lognormal distributions (18:6).

It appears that it was Galton who suggested the study of the lognormal when he pointed out that there are situations where the process of errors is multiplicative rather than additive as in normal theory (1:2). Galton explained that if $X1$, $X2$, ... Xn are n positive, independent random variables and

$$Tn = \sum_{i=1}^{n} Xi \qquad (10)$$

then

$$\log(Tn) = \sum_{i=1}^{n} \log(Xi) \qquad (11)$$

When one applies the Central Limit Theorem to the random variables, log(Xi), the resulting distribution of log(Tn) was basically the unit normal distribution as the sample size tends to infinity and as such, Tn was called the lognormal (18:7).

McAlister, in 1879, explicitly and in detail, set down the theory of the lognormal distribution (1:2). In his memoir presented to the Royal Society in London that year, he developed the expressions for mean, median, mode, and second moment of the lognormal along with the quartiles and octiles (1:2). According to Aitchison, the next "real" advance after McAlister's initial paper was that of Kapteryn in 1903. Kapteryn described a machine for generating lognormally distributed samples similar to that of Galton, used for normal or binormal samples (1:3).

Wiskell first the used method of moments to estimate parameters. He was also the first to consider that simple displacement of a variate rather than the variate itself is lognormally distributed (18:7). In this manner, the third parameter, the threshold parameter, was assigned to the value of the displacement, thus establishing the 3-LN distribution (1:4).

Applications

Aitchison found that the lognormal distribution can be used in the study of small particle statistics, economics, socialogy, biology, anthropometry, household size, physical and industrial processes, astronomy, and philology (1). The author notes that this list is in no way inclusive of all applications of the lognormal distribution. Examples of such processes include the distribution of personal incomes, inheritances and bank deposits, and the distribution of particle sizes (29:33).

While these are all important uses, current studies suggest the lognormal distribution will gain more importance with the Air Force's increased use of simulation models. "The log-normal distribution has been found to be applicable in describing time to failure for some types of components, and the literature seems to indicate increased use of this distribution in reliability models." (3:134)

Probability Distribution Function

In Edgeworth's "method of translation", one seeks a function of the observed random variable which closely approximates a random variable from the normal distribution. Johnson states that a variable, X, can be transformed to normality by a function, f(X). This function must be specialized and depends on a certain number of parameters (16:152). His transformation is as follows

$$z = \gamma + \delta \, f((X-\xi)/\lambda) \tag{12}$$

where

- f is a monotonic function of x and does not depend on any parameters
- z is the unit normal
- δ is a shape parameter
- γ is a shape parameter
- λ is the scale parameter
- ξ is the location parameter

The three-parameter lognormal probability density function (PDF) is derived by allowing the natural logarithmic function to be the function used in equation (12). By using natural logarithms, the scale parameter can be dropped (18:6). For this reason, all logrithms used in this thesis will be natural logrithms. By substituting these changes into equation (12), it becomes:

$$z = \gamma + \delta \ln(x - \xi) \tag{13}$$

When equation (13) is applied to the general form of the normal PDF, the PDF of the displaced lognormal variates follows:

$$F(X) = \frac{\delta}{\sqrt{2\pi}\,(X-\xi)} \exp\left[\frac{-(\gamma + \delta \ln(X-\xi))^2}{2}\right]$$

if $x > \xi$ (14)

$$F(X) = 0$$

if $x < \xi$

A more common expression of the lognormal PDF involves the mean (μ) and the standard deviation (σ^2) of the parent (original) normal distribution, where $\mu = -\gamma/\delta$ and $\sigma^2 = /\delta$ (18:9). By substituting these into equation (10), the new PDF is

$$F(X) = \frac{1}{\sigma\sqrt{2\pi}\,(X-\xi)} \left[\exp\left[\frac{-(\ln(X-\xi)-\mu)^2}{2\sigma^2}\right]\right.$$

if $x > \xi$ (15)

$$F(X) = 0$$

if $x < \xi$

Cumulative Distribution Function

The distribution functiton, also known as the cumulative distribution function (CDF), of a random variable, X, is the function that gives the probability that X is less than or equal to some number x (7:23). The CDF of a continuous random variable is found by intergrating its PDF over some given range (18:10).

The CDF, F(x), for the 3-LN is as follows (40:47)

$$F(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\xi}^{X} \frac{1}{u} \exp\left[\frac{-(\ln(u)-\mu)^2}{2\sigma^2} \right] du \quad (16)$$

where $u = X-\xi$.

IV. MAXIMUM LIKELIHOOD ESTIMATION

Introduction

Currently, the most used method of parameter estimation is the maximum likelihood method. Proposed by Daniel Bernolli in 1778, the concept of maximum likelihood was used by Gauss in developing his theory of least squares (18:22). According to Deutsch, maximum likelihood was not generally used as an estimation technique until 1912, when R.A. Fisher introduced a generalized recognized form. Fisher published a series of papers which extended Gauss' concepts to a comprehensive and unified system of mathematical statistics which has since had profound and wide development (9:135). Since then, the maximum likelihood method has been used successfully on most distributions. Maximum likelihood estimators (MLEs) have several very desirable properties. These include the fact that MLEs are consistent, asymptotically efficient and asympototically sufficient (25:167). These are properties of any good estimator: in addition, the MLE possesses a property of invariance (25:185). These properties will be discussed in more detail in a later section.

The principle of maximum likelihood consists in accepting as the best estimate of the parameters, say, $\theta_1, \theta_2, \ldots, \theta_K$, those values of the parameters which maximize the likelihood for a given set of observation, say, $x_1, x_2, \ldots, x_n$ (30:151). The population has a likelihood function, L defined as follows(18:23)

$$L = L(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_K)$$
$$= f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_K)$$
$$= f(x_i; \theta_1, \theta_2, \ldots, \theta_n)*f(x_i; \theta_1, \theta_2, \ldots, \theta_K), \ldots, f(x_i; \theta_1, \theta_2, \ldots, \theta_K)$$

$$= \quad f_i (x_i; \theta_1, \theta_2, \ldots, \theta_K) \tag{17}$$

If the population is discrete,

$$L(X_1, X_2, \ldots, X_n; \theta) = \prod_{i=1}^{n} p_i(\theta) \tag{18}$$

where $p_i(\theta)$ is the probability associated with the ith sample (9:135). Since MLEs are invariant, L and log(L) are maximized at the same values of θ_i. The log(L) yields a sum versus a product, which is more computationally efficient. For this reason, the logarithm of L is used in this thesis.

The likelihood function gives the "likelihood" that a set of random variables came from a certain density function (18:23). The values $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$, are the maximum likelihood estimators of the parameters $\theta_1, \theta_2, \ldots, \theta_K$. The process of maximizing the likelihood function is to take the partical derivative with respect to each parameter and set them to zero, then solve for the unknown parameters. This produces a system of k equations and k unknowns. These are

$$dL/d\theta_i = 0 \qquad i = 1, 2, \ldots k \tag{19}$$

Properties

As stated before, the properties of a maximum likelihood estimator (MLE) are those of being consistent, asymptotically efficient, asymptotically sufficient, and invariance. The concept of consistency is that, if for any two positive numbers, X and Y, there exists a number n_0 such at when n exceeds n_0, the probability

$$\left| t_n - \theta \right| > X \qquad (20)$$

is less than Y (30:151). This implies that as the sample size, n, increases the probability that the test statistic, t_n , and parameter, θ , will differ by any amount will decrease (30:151). This means as the sample size increases, the true value of the parameter will be approached. The estimator with the smallest asymptotic variance is called an efficient estimate (30:155). The estimator that converges the quickest to the true value of the parameter is preferred. The concept of the sufficient statistic, first developed by R.A. Fisher, states that a statistic that does exists and contains all the information about parameter that is in the sample is refered to as a suffient statistic (25:168). If a parameter, θ, has a MLE of $\hat{\theta}$, and "U(θ) is a function of θ with a single-valued inverse, the MLE of U(θ) is U($\hat{\theta}$)". (25:185)

Estimation of 3 Parameter Lognormal

In searching through literature on maximum likelihood estimation, few articles dealt with the 3 parameter lognormal distribution. Of those found a common thought prevailed: maximum likelihood estimation for the 3-LN is difficult at best.

By using methods previously discussed, the maximum likelihood equations for each of the three parameters of the 3-LN are (18:26)

$$\hat{\mu} = \sum_{i=1}^{n} \frac{\left[\ln(X_i - \xi)\right]}{n} \tag{21}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} \left[\ln(X_i - \xi) - \hat{\mu}\right]}{n}} \tag{22}$$

$$(\hat{\sigma} - \hat{\mu}) = \sum_{i=1}^{n} \frac{1}{X_i - \xi} \sum_{i=1}^{n} \frac{\ln(X_i - \xi)}{X_i - \xi} = 0 \tag{23}$$

According to Keefer (18), E. Wilson and J. Worchester first attempted to find the maximum likelihood estimator (MLE) of the three parameter lognormal in 1945 by using a trial and error method. This method proved to be "computationally ineffective" and resulted in "extremely poor parameter estimates" (18:27).

Next, in 1951, A. C. Cohen presented a more efficent and feasible technique for finding MLEs. He substituted equations (21) and (22) into equation (23). This produced a single function, f(ξ), with one unknown, ξ , the location parameter. Cohen then solved this equation using inverse interpolation over some small interval, say (ξ_1 , ξ_2), where $f(\xi) < 0$. The estimated value of the location parameter, $\hat{\xi}$, is then substituted into both equations (21) and (22), and solved, resulting in $\hat{\mu}$ and $\hat{\sigma}$.

Cohen also presented an alternate technique of estimation which is based on least observed value. The technique requires that

$$X_0 - \xi = \exp\left[\mu + \sigma * t_0\right] \tag{24}$$

where $X_0 = X_1 + (\tau/2)$. X_1 is the least observed sample value and τ, the the interval of precision or the smallest scale interval used in reading sample measurements (7:209). T_0 is determined from the relationship

$$\frac{k}{n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[\frac{-t^2}{2}\right] dt \tag{25}$$

where k is the number of times the least observed value occurs in the sample (6:209). By taking the natural logarithms of both sides, equation (24) becomes

$$\log (X_0 - \xi) = \left[\mu + \sigma * t_0\right] \tag{26}$$

Substituting equations (21) and (22) into equation (24), results in $F(\xi)$, a function of the location parameter, equivalent to the MLE of (18:28). A Monte Carlo analysis of the two techniques described by Cohen show that the method of least observed sample values provided better estimates than did the inverse interpolation method (18:28).

In 1963, B. M. Hill brought to light a significant fact regarding maximum likelihood estimation of the 3-LN distribution. He proved there exists a path referred to as the "path of no return" along which the likelihood function, $L(\xi, \hat{\mu}(\xi), \hat{\sigma}^2(\xi))$ of any sample, $X_1, X_2, \ldots, X_n$, tends to ∞ as ξ approaches X_1, the least observed value, and to a positive constant as ξ tends toward $-\infty$ (14:72). Allowing to converge to X_1 along this "path of no return", the estimates become unreasonable, namely $\hat{\xi} = X_1$, $\hat{\mu} = -\infty$, and $\hat{\sigma}^2 = +\infty$ (14:75). To avoid the problem of the "path of no return", Hill introduced a joint prior distribution for ξ, μ, and σ^2. By appling Bayes Theorem, this technique yielded the conclusion that the likelihood equations should be solved using $\hat{\xi}$, such that $\hat{\xi}$ satisfies equation (18:28)

$$\sum_{i=1}^{n} (X_i - \hat{\xi}) \frac{1}{\hat{\sigma}(\hat{\xi})} \sum_{i=1}^{n} \left[\frac{Z_j}{(X_i - \hat{\xi})} \right] = 0 \qquad (27)$$

where

$$Z_j = \frac{(\ln(X_i - \hat{\xi}) - \hat{\mu}(\hat{\xi})}{\hat{\sigma}(\hat{\xi})} \qquad (28)$$

In 1966, Dr H. Harter and Dr A. Moore published a paper reporting their method for 3-LN estimation. Recognizing the maximum likelihood equations for this distribution could not be solved algebraically and knowing that the likelihood function may take the "path of no return" and thus yield absurd estimates, they developed an iterative technique to solve the set of equations. The iterative process involves estimating the three parameters, one at a time in the cyclical order μ, σ, and ξ, omitting any assumed known parameters (18:30). To begin, the observed values are ordered and the initial estimates are chosen; for example, the initial estimate for ξ is X_1, the least observed value. Next, the iterative process begins; the false position (iterative linear interpolation) is used to determine the value, of the parameter being estimated, that satisfies the likelihood equation for that parameter. If no value of ξ satisfies the likelihood equation, the possiblity of encountering the "path of no return" occurs (13:848).

In 1973, C.J. Monlezun and L.A. Klinko, presented a paper at the thirty sixth annual Meeting of the Institute of Mathematical Statistics in New York City. In this paper, the author shows that "when the shape parameter for the lognormal distribution is assumed known, the likelihood equations have a unique solution at which the likelihood function attains its maximum value" (24:2). For this reason, this paper is used as the basis for the calculations in this thesis for estimating the MLEs of the 3-LN distribution.

A random variable X is said to be lognormally distributed for some constant ξ, less than X; the log $(X-\xi)$ has a normal distribution with mean, μ, and variance, σ^2, and the density of X is then

$$f(X; \xi, \mu, \sigma^2) = \frac{\sqrt{(2\pi\sigma^2)}}{(X-\xi)} * \exp\left[\frac{-(\ln(X-\xi)-\mu)^2}{2\sigma^2}\right] \qquad (29)$$

for $X > \xi$. By letting $L(\xi, \mu, \sigma^2)$ denote the likelihood function of n independent observation of X.

As stated in chapter III, the CDF of the lognormal is as follows

$$F(X) = \frac{1}{(X-\xi)\sigma\sqrt{2\pi}} \exp\left[\frac{-\ln((X-\xi)/\mu)^2}{2\sigma^2}\right] \tag{30}$$

by setting $\phi = \ln \mu$,

$$F(X) = \frac{1}{((X-\xi)\sigma\sqrt{2\pi})} \exp\left[\frac{\ln((X-\xi)-\phi)^2}{-2\sigma^2}\right] \tag{31}$$

the likelihood function, L, is

$$L = \prod_{i=1}^{n}\left[\frac{1}{((X_i-\xi)\sigma\sqrt{2\pi})} \exp\left[-2\sigma^2(\ln(X-\xi)-\phi)^2\right]\right] \tag{32}$$

$$\operatorname{Ln} L = \sum_{i=1}^{n}\left[\ln\left[\frac{1}{(X_i-\xi)\sigma\sqrt{2\pi}}\right] + \left[\frac{((\ln(X_i-\xi)-\phi))^2}{-2\sigma^2}\right]\right] \tag{33}$$

$$\operatorname{Ln} L = \sum_{i=1}^{n}\left[-\ln(X_i-\xi) - \ln(\sigma\sqrt{2\pi}) + \frac{(\ln(X_i-\xi)-\phi)^2}{(-2\sigma^2)}\right] \tag{34}$$

By setting the partial deriviatives to zero, the partial with respect to the scale parameter is as follows

$$dLnL/d\mu = 0 \tag{35}$$

results in the equation

$$\hat{\mu} = \hat{\mu}(\xi) = 1/n * \sum_{i=1}^{n} \log(X_i + \xi) \qquad (36)$$

and the partial with respect to the shape parameter yields the following equation:

$$dL/d\sigma^2 = 0 \qquad (37)$$

results in the equation

$$\hat{\sigma}^2 = \hat{\sigma}^2(\xi) = 1/n * \sum_{i=1}^{n} (\log(X_i + \xi) - \hat{\mu}(\xi)) \qquad (38)$$

For ξ fixed, $L(\xi, \mu, \sigma^2)$ reaches a maximum at $(\xi, \hat{\mu}(\xi), \hat{\sigma}^2(\xi))$. Monlezun shows that for a known shape parameter σ^2, the equation

$$dL(\xi)/d\xi = 0 \qquad (39)$$

has a unique solution, say $\xi = \xi$, that satisfies the equation,

$$\frac{1}{(X_i - \xi)} + \frac{1}{(\sigma^2 * (X + \xi))} * (\ln(X - \xi) - u) = 0 \qquad (40)$$

where $u = 1/n \sum_{i=1}^{n} \ln(X - \xi)$.

V. METHODOLOGY

As stated in chapter I, there were three major phases in this research effort. The first step in developing the modified goodness-of-fit test was to construct tables of critical values by Monte Carlo method. This method was first used by Lilliefors (1967) in his research of fitting the normal distribution when the mean and variance were unknown (36). The next phase was to compare the powers of the modified goodness-of-fit tests. Finally, the relationship between the shape parameter and the critical values generated was determined.

For the first phase, each test procedure was modified by generating random deviates which followed a lognormal distribution. The random deviates were then ordered in ascending order. These ordered deviates were used to estimate the scale and location parameters using maximum likelihood estimation (MLE). The next step in the first phase was the estimate parameters from the n ordered lognormal deviates and use these estimates to calculate the hypothesized distribution function (33). Each of the above steps was repeated 5000 times for each of the statistical values being tested. The final step was to arrange the critical values into tabular form for ease in reading and use of the test statistics.

The second phase of the research, comparing the powers of the tests involved, tested the null hypothesis that the data was from a lognormal distribution. The proportion of the time that the test rejected this null hypothesis was counted for each sample taken. The power was the percentage that the test rejects the null hypothesis.

The final phase of the research was to determine the relationship, if any, between the shape parameter and the critical values. This functional formula can be used to interpolate any values not found in the tables.

The Monte Carlo Method

"Mathematics can be divided into theoretical and experimential categories." (28:4-1) The primary difference between the two is that theoreticians deduce conclusions from postulates, experimentalists arrive at conclusions from observations (12:1). The Monte Carlo method is a branch of experimental mathematics where random numbers are generated to provide data for these experiments to simulate observations. This method is often used in fields where real world data is expensive or even impossible to obtain; for example, when studying nuclear effects.

Identifying Critical Values

Each group of n lognormal deviates represent a sample of size n from a lognormal destribution with known parameters. For this reason, the null hypothesis "H_0: H(X) = lognormal CDF" is true for each sample. Using the K-S, the A-D, and the C-VM tests for goodness of fit, 5000 independent test statistics were calculated using a known CDF for each test. The 5000 test statistics for each test were then arranged in ascending order using an IMSL (15) subroutine, VSRTA . The next step was to identify the "critical region", that is, where the test statistic would wrongly reject the known true null hypothesis (18:4-10). Next, the critical values are selected according to desired "level of significance", or α , which is the maximum probability of rejecting a true null hypothesis.

Since H_0 is true for all the calculated test statistics, and α is the maximum probability of rejecting H_0, then $1-\alpha$ is the minimum probability of correctly accepting the null hypothesis. The value, $1-\alpha$, is the percentage of total test statistics within the critical region. For example, the 95th percentile is a number that the test statistic will exceed 5% of the time or less and will be less than with probability of .95 or less (7:39). Using this system of percentages, the critical values were determined from the 5000 test statistics.

In the first phase, generating critical value tables, a FORTRAN program, written by Porter (28), was adapted to perform the Monte Carlo simulation necessary of this research objective. The flow chart and code for this program is located in appendix A.

The nine steps followed in this thesis to accomplish this are as follow (28:4-19-4-21).

Step 1 - Generate the data. In this thesis, sample observations were generated by a computer program avaliable in the International Mathematics Statistics Library (IMSL). This subroutine, GGLNG, generated lognormal random deviates from a two-parameter lognormal distribution, to which a location parameter of say, 10, was added.

Step 2 - Order the data. The random deviates were arranged in asending order using an IMSL subroutine, VSRTA.

Step 3 - Estimate the parameters. The maximum likelihood estimators of the scale and location parameters were found using the method described in chapter I.

Step 4 - Compute hypothesized CDF. Using the estimated parameters, found in step 3, and the ordered sample generated in step 2, the hypothesized CDF is calculated.

Step 5 - Calculate the test statistics. The modified test statistics are calculated using equations (2), (4), and (6).

Step 6 - Generate 5000 test statistics. Repeat steps 1 thru 5, 5000 times. This is necessary for the Monte Carlo simulation. This generates 5000 independent test statistics for each of the three tests, K-S, A-D, and C-VM.

Step 7 - Determine the critical values. The 5000 test statistics generated in step 6 are ordered using the IMSL subroutine, VSRTA, Determine the 80th, 85th, 90th, 95th, and 99th percentile of the 5000 statistics, these correspond to the .20, .15, .10, .05, and .01 levels of significance. That is the 4000th test statistic is 80% of the 5000 statistics; therefore, it becomes the critical value for a significance level of .20.

Step 8 - Repeat for sample size. To study the effect of the sample size on critical values, repeat step 1 thru step 7 for sample size n where n = 5, 10, 15, 20, 25, and 30

Step 9 - Repeat for shape parameters. The known shape parameter ranged from 1 to 4 in steps of .5.

The critical values calculated are found in tables I, II and III, found in the chapter that follows.

Comparing Powers

As explained earlier, the probability of correctly rejecting a false null hypothesis is known as the power of the test; therefore, the higher the power, the more useful the test. In this thesis, the null hypothesis was that the random deviates being tested follow a lognormal distribution with the shape parameter known (in this thesis, shape = 1 and 3). The alternate hypothesis was that, the deviates followed a distribution other than the lognormal. A FORTRAN program, written by Porter (28), was adapted to perform the power comparison necessary. The flow chart and code for this program are found in Appendix B.

Step 1 - Generate the data. Random deviates, for sample size n, were generated using the IMSL subroutines GGWIB, GGAMR, GGBTR, GGEXN, and GGNML. These alternate distributions were the Weibull with shape of 3.5, the gamma with shape of 2.0, the beta with parameters of p=2 and q=3, and the normal distribution. Two sets of lognormal deviates were also tested. The first with the shape of 1.0, the second with shape of 3.0.

Step 2 - Order the random deviates. The subroutine VSRTA from IMSL was used to arrange the data in ascending order.

Step 3 - Estimate the parameters. Using the technique of maximum likelihood estimation, described in chapter 4, estimate the scale and location parameters.

Step 4 - Compute the hypothesized distribution function. Using the estimated parameters, found in step 3, and the ordered sample generated in step 2, the subroutine HYPCDF calculates the hypothesized CDF.

Step 5 - Calculate the modified K-S, A-D, and C-VM test statistics. The modified test statistics are caluclated by the subroutine TESTAT, using the equations (2), (4), and (6).

Step 6 - Repeat 5000 times. The Monte Carlo simulation of observational data uses 5000 repeatitions.

Step 7 - Determine the power of the test. By counting the number of times the null hypothesis is rejected and divided by 5000. This is the power of the test.

Step 8 - Repeat for alternate distributions. Repeat steps 1-7 for each of the alternate distributions; that is, the Weibull, Gamma, Beta, Exponential, and the Normal.

Step 9 - Repeat for sample size. Repeat steps 1-8 for the sample sizes, n=5, 15, and 25.

Step 10 - Repeat for levels of significance. Repeat steps 1-9 for the levels of significance used in this thesis; that is, α = .05 and .01.

Step 11 - Repeat for shape parameters. Repeat steps 1-10 for the Lognormal with shape of 3.0 (in the first replication, shape = 1.0).

The results of the power comparison study are found in tables IV and V.

Determining Functional Relationship

The final stage of this thesis research was determing the functional relationship, if any, between the known shape parameters and the modified critical value. This relationship can be used to find values not located in the tables generated in this thesis.

This phase of the research was completed using SAS (32) to preform a quadratic regression. The model used is as follows:

$$Y = B_0 + B_1 X + B_2 X{*}{*}2 \tag{41}$$

where

Y = critical value
X = shape parameter

The results of these linear regressions are located in Tables VI, VII, and VIII.

VI. RESULTS

The results of this thesis are such that each research objective listed in Chapter I has been successfully completed. Tables containing the Modified Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Cramer-von Mises (C-VM) critical values have been generated. The critical values are documented in Tables I, II, and III. The power of each of these new tests have been tested using five alternated distributions and the lognormal (with shape = 1.0 and 3.0). These values are documented in Tables IV and V. The third objective has been completed by the generation of Tables VI, VII, and VIII, showing the coefficents and the correlation value, R^2 (which indicates the percent of total variation explained by the regression curve).

Critical Value Tables

Table I contains the critical values for the modified Kolmogorov-Smirnov test. The new Anderson-Darling statistics are found in Table II. Cramor-von Mises critical values are located in Table III. Each table includes the test statistic generated with sample sizes of 5, 10, 15, 20, 25, and 30. The shape parameters ranged from 1 to 4 in increments of .5. The levels of signifiance used were .20, .15, .10, .05, and .01.

TABLE I
CRITICAL VALUES FOR THE MODIFIED K-S TEST

ALPHA	N	C=1.0	1.5	2.0	2.5	3.0	3.5	4.0
.20	5	0.307	0.292	0.296	0.308	0.322	0.333	0.343
.20	10	0.226	0.230	0.261	0.288	0.309	0.328	0.344
.20	15	0.188	0.204	0.242	0.274	0.297	0.320	0.338
.20	20	0.163	0.190	0.233	0.266	0.292	0.313	0.331
.20	25	0.148	0.179	0.223	0.261	0.286	0.309	0.328
.20	30	0.135	0.170	0.218	0.256	0.282	0.306	0.325
.15	5	0.319	0.306	0.306	0.318	0.332	0.341	0.350
.15	10	0.236	0.240	0.271	0.298	0.319	0.336	0.350
.15	15	0.196	0.213	0.251	0.283	0.305	0.327	0.343
.15	20	0.171	0.197	0.241	0.274	0.299	0.320	0.337
.15	25	0.154	0.187	0.230	0.267	0.293	0.315	0.333
.15	30	0.141	0.178	0.224	0.261	0.288	0.311	0.329
.10	5	0.337	0.322	0.320	0.330	0.343	0.351	0.358
.10	10	0.251	0.254	0.284	0.309	0.329	0.346	0.359
.10	15	0.207	0.224	0.263	0.294	0.315	0.336	0.351
.10	20	0.182	0.208	0.250	0.282	0.307	0.328	0.344
.10	25	0.162	0.196	0.240	0.276	0.301	0.322	0.340
.10	30	0.150	0.187	0.232	0.270	0.295	0.318	0.336
.05	5	0.364	0.345	0.343	0.348	0.358	0.363	0.370
.05	10	0.271	0.276	0.303	0.326	0.343	0.359	0.371
.05	15	0.226	0.241	0.279	0.308	0.328	0.348	0.362
.05	20	0.195	0.222	0.263	0.297	0.319	0.339	0.354
.05	25	0.177	0.211	0.254	0.288	0.313	0.332	0.350
.05	30	0.163	0.201	0.244	0.283	0.306	0.327	0.345
.01	5	0.413	0.387	0.381	0.378	0.380	0.384	0.390
.01	10	0.312	0.317	0.336	0.355	0.370	0.381	0.392
.01	15	0.262	0.272	0.309	0.336	0.352	0.370	0.384
.01	20	0.225	0.255	0.289	0.321	0.342	0.360	0.375
.01	25	0.204	0.241	0.284	0.311	0.333	0.353	0.367
.01	30	0.186	0.229	0.268	0.304	0.327	0.347	0.365

TABLE II
CRITICAL VALUES FOR THE MODIFIED A-D TEST

ALPHA	N	C=1.0	1.5	2.0	2.5	3.0	3.5	4.0
.20	5	0.606	0.455	0.474	0.533	0.604	0.665	0.728
.20	10	0.588	0.573	0.825	1.079	1.279	1.463	1.641
.20	15	0.599	0.732	1.191	1.638	1.963	2.279	2.564
.20	20	0.595	0.910	1.586	2.198	2.686	3.122	3.515
.20	25	0.593	1.055	1.937	2.775	3.390	3.960	4.480
.20	30	0.598	1.223	2.315	3.352	4.093	4.814	5.408
.15	5	0.660	0.493	0.515	0.572	0.643	0.700	0.761
.15	10	0.652	0.637	0.901	1.148	1.354	1.529	1.700
.15	15	0.663	0.805	1.294	1.725	2.051	2.375	2.652
.15	20	0.657	1.002	1.706	2.310	2.804	3.228	3.607
.15	25	0.654	1.156	2.057	2.904	3.518	4.088	4.583
.15	30	0.659	1.350	2.466	3.502	4.238	4.936	5.532
.10	5	0.742	0.547	0.562	0.626	0.692	0.749	0.801
.10	10	0.744	0.726	0.993	1.240	1.443	1.616	1.767
.10	15	0.739	0.910	1.411	1.855	2.179	2.483	2.753
.10	20	0.739	1.127	1.842	2.447	2.941	3.372	3.727
.10	25	0.744	1.287	2.225	3.074	3.681	4.246	4.717
.10	30	0.750	1.511	2.638	3.698	4.431	5.118	5.688
.05	5	0.904	0.636	0.648	0.705	0.766	0.810	0.863
.05	10	0.899	0.860	1.143	1.389	1.586	1.736	1.886
.05	15	0.900	1.052	1.598	2.030	2.372	2.641	2.922
.05	20	0.906	1.302	2.056	2.678	3.156	3.563	3.910
.05	25	0.898	1.524	2.497	3.337	3.941	4.463	4.924
.05	30	0.891	1.759	2.927	3.974	4.688	5.394	5.932
.01	5	1.279	0.816	0.796	0.839	0.882	0.928	0.963
.01	10	1.247	1.162	1.413	1.638	1.826	1.941	2.092
.01	15	1.296	1.379	1.962	2.457	2.708	2.947	3.201
.01	20	1.274	1.788	2.452	3.135	3.514	3.952	4.219
.01	25	1.242	2.006	3.019	3.798	4.364	4.876	5.330
.01	30	1.244	2.333	3.450	4.520	5.213	5.830	6.404

TABLE III
CRITICAL VALUES FOR THE MODIFIED C-VM TEST

ALPHA	N	C=1.0	1.5	2.0	2.5	3.0	3.5	4.0
.20	5	0.091	0.080	0.086	0.099	0.113	0.126	0.138
.20	10	0.091	0.100	0.149	0.199	0.240	0.280	0.316
.20	15	0.092	0.126	0.210	0.297	0.365	0.433	0.493
.20	20	0.091	0.153	0.277	0.396	0.499	0.590	0.673
.20	25	0.091	0.175	0.334	0.498	0.626	0.746	0.856
.20	30	0.091	0.202	0.396	0.599	0.757	0.906	1.032
.15	5	0.100	0.087	0.094	0.107	0.122	0.134	0.146
.15	10	0.101	0.114	0.164	0.214	0.257	0.296	0.331
.15	15	0.103	0.141	0.231	0.318	0.385	0.455	0.512
.15	20	0.102	0.171	0.301	0.421	0.525	0.616	0 695
.15	25	0.102	0.195	0.359	0.524	0.656	0.776	0.881
.15	30	0.102	0.226	0.425	0.631	0.786	0.934	1.061
.10	5	0.113	0.098	0.106	0.119	0.134	0.146	0.155
.10	10	0.115	0.132	0.185	0.233	0.278	0.316	0.347
.10	15	0.117	0.162	0.256	0.345	0.413	0.480	0.536
.10	20	0.117	0.195	0.331	0.451	0.556	0.649	0.724
.10	25	0.116	0.221	0.393	0.562	0.693	0.813	0.912
.10	30	0.115	0.253	0.463	0.674	0.830	0.975	1.097
.05	5	0.133	0.117	0.125	0.137	0.151	0.160	0.170
.05	10	0.141	0.161	0.218	0.269	0.310	0.343	0.375
.05	15	0.142	0.190	0.298	0.385	0.458	0.518	0.577
.05	20	0.142	0.232	0.374	0.502	0.604	0.693	0.765
.05	25	0.141	0.266	0.452	0.621	0.749	0.862	0.961
.05	30	0.141	0.302	0.522	0.734	0.886	1.037	1.153
.01	5	0.172	0.154	0.159	0.170	0.178	0.187	0.193
.01	10	0.197	0.223	0.277	0.325	0.366	0.394	0.425
.01	15	0.199	0.258	0.374	0.479	0.534	0.588	0.643
.01	20	0.202	0.323	0.456	0.603	0.686	0.781	0.840
.01	25	0.204	0.358	0.559	0.723	0.844	0.956	1.058
.01	30	0.199	0.414	0.636	0.854	1.007	1.139	1.263

Power Comparison Tables

The results from the power comparison program are found in tables IV and V. For each table, the sample size varied as n = 5, 15, and 25 and the power comparisons are shown at the significance levels of .05 and .01.

The first column of the power comparison tables is approximately the level of significance since the null hypothesis, that the observed sample came from a lognormal distribution, is true. In Table IV, the data came the lognormal with shape of 1; in Table V, the first column contained data from a lognormal distribution with shape of 3. In the last five columns of the tables, the data did come from five different distributions and the values shown under these headings are the respective powers against accepting the data as lognormally distributed, given that it came from the respective distributions. The alternate distributions included in the power comparison were the Weibull with shape of 3.5, the gamma with shape of 2, the beta with p=2 and q=3, the exponential with mean of 2 and the normal distribution.

Regression Tables

Tables VI, VII, and VIII document the relationship between the modified critical values for the three different tests and the shape parameter. These tables can be used to find critical values not included in the tables generated (Tables VI, VII, and VIII). For observed sample size of 10, 15, 20, 25, and 30, with a shape parameter betweem 1.0 and 4.0, and a level of significance of .20, .15, .10, .05, and

TABLE IV
POWER TEST FOR THE LOGNORMAL DISTRIBUTION

H_O: LOGNORMAL DISTRIBUTION AT SHAPE C = 1.0
H_A: THE DATA FOLLOW ANOTHER DISTRIBUTION

LEVEL OF SIGNIFICANCE = .05

		ALTERNATE DISTRIBUTIONS					
N	TEST	LOG.1	WEIBL	GAMMA	BETA	EXPON	NORML
5	K-S	0.050	0.130	0.054	0.102	0.056	0.131
5	A-D	0.050	0.026	0.016	0.026	0.036	0.021
5	CVM	0.050	0.111	0.047	0.091	0.052	0.109
15	K-S	0.041	0.589	0.128	0.343	0.065	0.600
15	A-D	0.044	0.665	0.108	0.387	0.051	0.680
15	CVM	0.041	0.719	0.148	0.448	0.064	0.728
25	K-S	0.052	0.880	0.222	0.633	0.091	0.892
25	A-D	0.053	0.952	0.249	0.777	0.072	0.952
25	CVM	0.055	0.952	0.286	0.780	0.089	0.956

LEVEL OF SIGNIFICANCE = .01

		ALTERNATE DISTRIBUTIONS					
N	TEST	LOG.1	WEIBL	GAMMA	BETA	EXPON	NORML
5	K-S	0.008	0.025	0.007	0.015	0.008	0.023
5	A-D	0.010	0.000	0.003	0.003	0.006	0.000
5	CVM	0.010	0.024	0.010	0.016	0.010	0.025
15	K-S	0.010	0.303	0.036	0.130	0.016	0.341
15	A-D	0.011	0.358	0.022	0.135	0.006	0.400
15	CVM	0.010	0.476	0.046	0.221	0.016	0.527
25	K-S	0.009	0.703	0.077	0.349	0.018	0.729
25	A-D	0.013	0.840	0.091	0.514	0.012	0.863
25	CVM	0.012	0.842	0.110	0.531	0.016	0.873

NOTE: Since H_O is true, the LOG.1 column contains the level of significance

TABLE V
POWER TEST FOR THE LOGNORMAL DISTRIBUTION

H_O: LOGNORMAL DISTRIBUTION AT SHAPE C = 3.0
H_A: THE DATA FOLLOW ANOTHER DISTRIBUTION

LEVEL OF SIGNIFICANCE = .05

		ALTERNATE DISTRIBUTIONS					
N	TEST	LOG.3	WEIBL	GAMMA	BETA	EXPON	NORML
5	K-S	0.055	0.229	0.091	0.189	0.052	0.221
5	A-D	0.052	0.267	0.104	0.229	0.051	0.260
5	CVM	0.053	0.256	0.105	0.209	0.052	0.260
15	K-S	0.065	0.879	0.285	0.749	0.080	0.861
15	A-D	0.058	0.813	0.287	0.632	0.070	0.814
15	CVM	0.058	0.805	0.281	0.608	0.068	0.808
25	K-S	0.041	0.990	0.419	0.962	0.061	0.980
25	A-D	0.044	0.960	0.424	0.844	0.053	0.951
25	CVM	0.043	0.958	0.415	0.835	0.052	0.955

LEVEL OF SIGNIFICANCE = .01

		ALTERNATE DISTRIBUTIONS					
N	TEST	LOG.3	WEIBL	GAMMA	BETA	EXPON	NORML
5	K-S	0.011	0.074	0.020	0.047	0.011	0.063
5	A-D	0.010	0.074	0.025	0.066	0.013	0.072
5	CVM	0.011	0.077	0.026	0.060	0.014	0.078
15	K-S	0.016	0.663	0.095	0.425	0.014	0.684
15	A-D	0.012	0.585	0.095	0.340	0.013	0.623
15	CVM	0.012	0.584	0.096	0.332	0.013	0.625
25	K-S	0.012	0.942	0.194	0.815	0.017	0.929
25	A-D	0.011	0.873	0.211	0.623	0.013	0.883
25	CVM	0.011	0.870	0.211	0.620	0.012	0.888

NOTE: Since H_O is true, the LOG.3 column contains the level of significance

Table VI

COEFFICIENT AND R^2 VALUES OF THE RELATIONSHIP
BETWEEN KOLMOGOROV-SMIRNOV CRITCAL VALUES
AND LOGNORMAL SHAPE PARAMETERS
1.0 < shape < 4.0

		LEVEL OF SIGNIFICANCE				
n	Coeff	.20	.15	.10	.05	.01
10	b0	.1739	.1809	.1995	.2229	.2756
	b1	.0456	.0492	.0455	.0438	.0330
	b2	-.0006	-.0015	-.0012	-.0015	-.0009
	R^2	.9837	.9813	.9806	.9844	.9864
15	b0	.1138	.1186	.1279	.1508	.1941
	b1	.0724	.0766	.0791	.0748	.0663
	b2	-.0040	-.0050	-.0057	-.0054	-.0046
	R^2	.9925	.9923	.9918	.9907	.9857
20	b0	.0687	.0755	.0895	.0999	.1380
	b1	.0980	.0992	.0962	.0998	.0931
	b2	-.0080	-.0084	-.0080	-.0090	-.0084
	R^2	.9963	.9955	.9961	.9959	.9983
25	b0	.0459	.0508	.0569	.0741	.1029
	b1	.1071	.1092	.1122	.1105	.1116
	b2	-.0091	-.0096	-.0103	-.0104	-.0114
	R^2	.9968	.9977	.9978	.9981	.9986
30	b0	.0239	.0303	.0400	.0529	.0794
	b1	.1180	.1186	.1181	.1190	.1180
	b2	-.0106	-.0110	-.0110	-.0115	-.0117
	R^2	.9974	.9984	.9982	.9981	.9994

Relationship between the critical values Y and the shape Parameter X, is:

$$Y = b0 + b1\ X + b2\ X \quad \text{where } 1.0 < X < 4.0$$

Table VII

COEFFICIENT AND R^2 VALUES OF THE RELATIONSHIP BETWEEN ANDERSON-DARLING CRITCAL VALUES AND LOGNORMAL SHAPE PARAMETERS

1.0 < shape < 4.0

		LEVEL OF SIGNIFICANCE				
n	Coeff	.20	.15	.10	.05	.01
10	b0	.2302	.2714	.3439	.4818	.9189
	b1	.2621	.2884	.3102	.3250	.2204
	b2	.0246	.0192	.0137	.0087	.0203
	R^2	.9798	.9787	.9762	.9697	.9566
15	b0	-.2262	-.1993	-.1989	-.0882	.2266
	b1	.7446	.7903	.8850	.9346	1.0076
	b2	-.0095	-.0171	-.0348	-.0438	-.0644
	R^2	.9898	.9902	.9906	.9883	.9771
20	b0	-.7278	-.7462	-.7369	-.6701	-.3786
	b1	1.2722	1.3738	1.4700	1.5853	1.7239
	b2	-.0504	-.0690	-.0860	-.1080	-.1416
	R^2	.9940	.9944	.9952	.9945	.9962
25	b0	-1.2097	-1.2433	-1.2690	-1.2689	-1.0561
	b1	1.7603	1.8785	2.0231	2.2314	2.4490
	b2	-.0817	-.1027	-.1290	-.1691	-.2134
	R^2	.9947	.9950	.9953	.9963	.9980
30	b0	-1.7126	-1.7585	-1.8011	-1.8176	-1.6433
	b1	2.2902	2.4366	2.6084	2.8267	3.1191
	b2	-.1239	-.1507	-.1814	-.2203	-.2776
	R^2	.9953	.9961	.9965	.9976	.9990

Relationship between the critical values Y and the shape Parameter X, is:

$$Y = b0 + b1\ X + b2\ X \quad \text{where } 1.0 < X < 4.0$$

Table VIII

COEFFICIENT AND R^2 VALUES OF THE RELATIONSHIP BETWEEN CRAMER-VON MISES CRITCAL VALUES AND LOGNORMAL SHAPE PARAMETERS
1.0 < shape < 4.0

		LEVEL OF SIGNIFICANCE				
n	Coeff	.20	.15	.10	.05	.01
10	b0	.0134	.0164	.0201	.0328	.0864
	b1	.0633	.0722	.0853	.1018	.1080
	b2	.0034	.0020	-.0005	-.0038	-.0057
	R^2	.9889	.9903	.9909	.9907	.9932
15	b0	-.0611	-.0636	-.0667	-.0651	-.0525
	b1	.1380	.1542	.1755	.2022	.2561
	b2	.0006	-.0021	-.0058	-.0101	-.0205
	R^2	.9934	.9937	.9944	.9937	.9915
20	b0	-.1478	-.1564	-.1606	-.1681	-.1476
	b1	.2249	.2488	.2729	.3127	.3681
	b2	-.0044	-.0084	-.0124	-.0194	-.0300
	R^2	.9951	.9954	.9961	.9964	.9974
25	b0	-.2250	-.2354	-.2529	-.2706	-.2480
	b1	.2990	.3250	.3637	.4195	.4771
	b2	-.0065	-.0108	-.0174	-.0275	-.0377
	R^2	.9951	.9955	.9959	.9970	.9983
30	b0	-.3110	-.3216	-.3486	-.3600	-.3665
	b1	.3851	.4139	.4624	.5121	.6045
	b2	-.0114	-.0163	.0245	-.0329	-.0493
	R^2	.9954	.9961	.9964	.9975	.9991

Relationship between the critical values Y and the shape Parameter X, is:

$$Y = b0 + b1\,X + b2\,X \quad \text{where } 1.0 < X < 4.0$$

.01, the coefficients, B_0, B_1, and B_2, found in Table VI can be substituted into the equation $Y=B_0+B_1*X + B_2*(X**2)$ to find the K-S critical values not found in Table I. Tables VII and VIII can be used similarily for the A-D and the C-VM critical values, respectively. These regression tables also contain the R^2 value which indicates the percent of total variation explained by the regression. This means, the closer the R^2 value is to 1, the stronger the regression model is in calculating the additional critical values.

Recommendations

This thesis is the latest in a series of research done on modified goodness-of-fit statistics for various distributions. Follow on study could be varing the parameter of the program that generated the critical values. By rerunning the current program with different sample sized or shape parameters, the effect of larger sample sizes on the modified tests can be studied.

Other branches to investagate could include using estimators other than the Maximum Likelihood Estimators for parameter estimation. By increasing the sample size, the Chi-square may be brought more into consideration while comparing the power of the other three tests.

This type of research can become increasingly useful with the current trend of increased use of simulation models in both the private and military arena. Since real world data seldom have known parameters, this non-parametric testing will become more nd more helpful in modeling real world events.

APPENDIX A

Flow Chart for Program Critical

Computer Program and Subroutines
for Generating Critical Value
Tables for Modifed Goodness-of-Fit
Tests

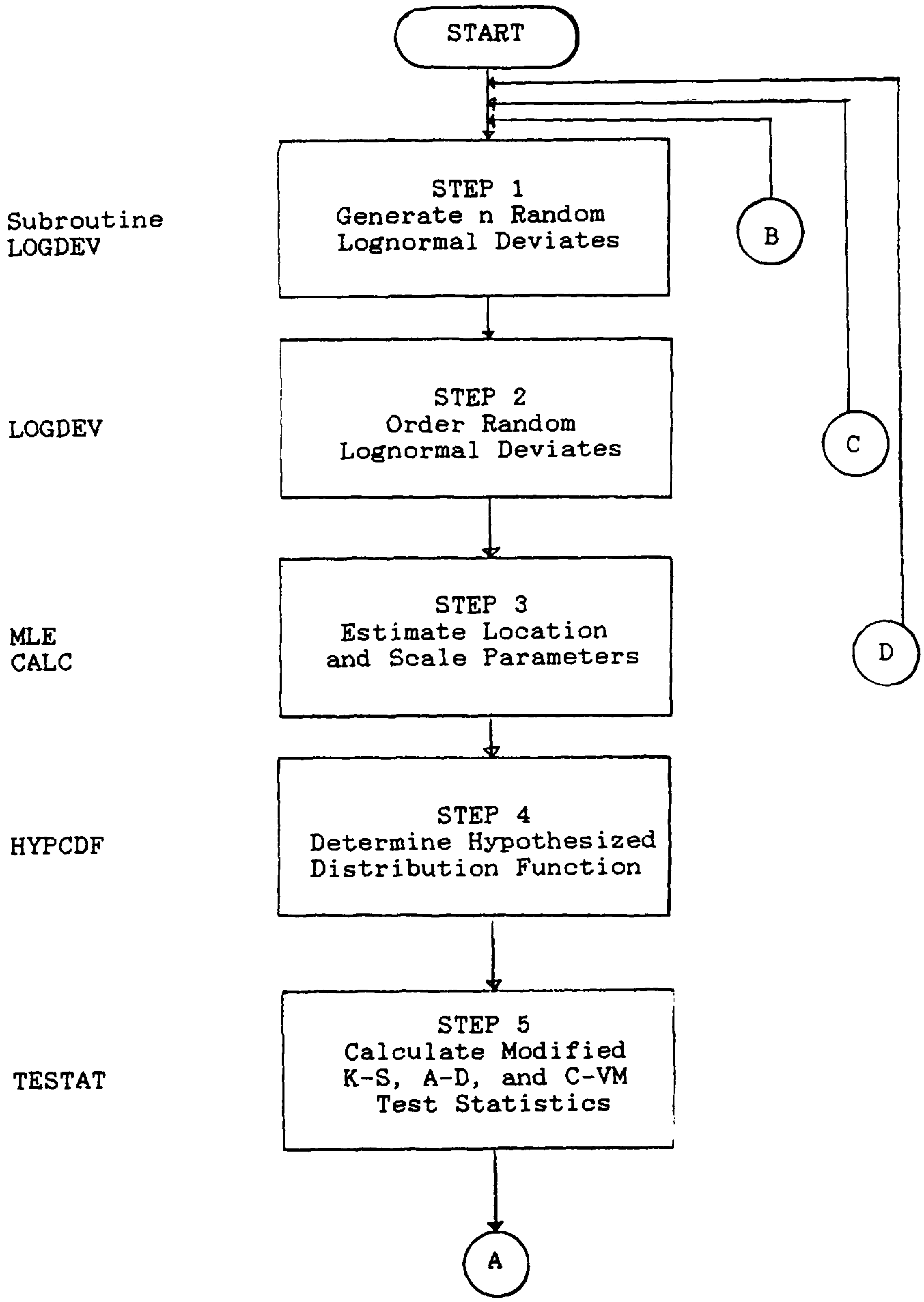


Fig 1. Flow chart for Program CRITICAL

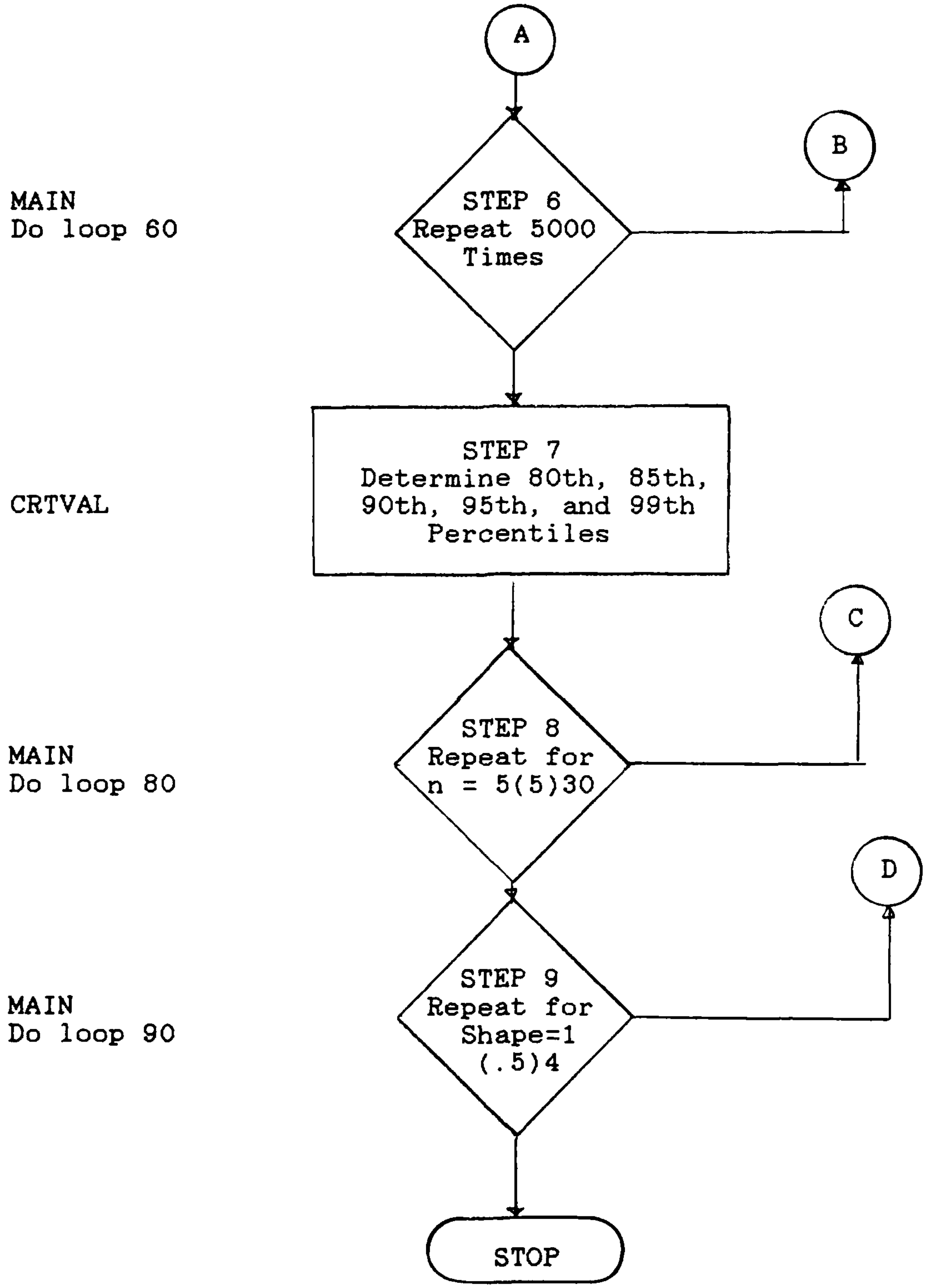


Fig 1 (Continued). Flow Chart for Program Critical

```
***************************************************************
***************************************************************
*                        PROGRAM CRITICAL
***************************************************************
***************************************************************
*
*  PROGRAM FOR LOGNORMAL GOODNESS-OF-FIT TESTS
*
*  PURPOSE:  1. GENERATE CRITICAL VALUE TABLES FOR THE MODIFIED
*               K-S, A-D, AND C-VM TESTS FOR THE THREE-
*               PARAMETER LOGNORMAL DISTRIBUTION WHEN LOCATION
*               AND SCALE PARAMETERS MUST BE ESTIMATED FROM THE
*               SAMPLE
*
*  VARIABLES:
*     DSEED = RANDOM NUMBER SEED
*         C = SHAPE PARAMETER
*         X = ARRAY OF LOGNORMAL RANDOM DEVIATES
*         N = SAMPLE SIZE
*        NC = SAMPLE SIZE * SHAPE PARAMETER
*      AMLE = MLE OF THE LOCATION PARAMETER
*      BMLE = MLE OF THE SCALE PARAMETER
*         P = ARRAY OF N POINTS OF HYPOTHESIZED CDF
*       PCT = PERCENTILE VALUE
*        KS = ARRAY OF VALUES OF MOD. K-S TEST STATISTIC
*        AD = ARRAY OF VALUES OF MOD. A-D TEST STATISTIC
*       CVM = ARRAY OF VALUES OF MOD. C-VM TEST STATISTIC
*        IT = ITERATION COUNTER (5000 REQ. FOR MONTE CARLO)
*      NSIZ = SAMPLE SIZE COUNTER
*      NSHP = SHAPE PARAMETER COUNTER
*      NPCT = PERCENTILE COUNTER
*       NST = NUMBER OF REPETITIONS
*    KSCRIT = ARRAY OF K-S CRITICAL VALUES
*    ADCRIT = ARRAY OF A-D CRITICAL VALUES
*    CVCRIT = ARRAY OF C-VM CRITICAL VALUES
*         Y = ARRAY OF PLOTTING POSITIONS
*     ALPHA = LEVEL OF SIGNIFICANCE
*
*  INPUTS:
*              NST = NUMBER OF REPETITIONS
*            DSEED = RANDOM NUMBER SEED
*
*  SUBROUTINES:
*      LOGDEV - GENERATES N ORDERED LOGNORMAL DEVIATES
*        FILL - ZEROS ALL ARRAYS
*         MLE - CALCULATES MAXIMUM LIKELIHOOD ESTIMATORS
*        CALC - PERFORMS NECESSARY CALCULATION FOR MLE
*      HYPCDF - COMPUTES THE HYPOTHESIZED LOGNORMAL CDF
*      TESTAT - CALCULATES THE K-S, A-D, C-VM TEST STATISTICS
*      CRTVAL - DETERMINES CRIT. VALUES FROM PLOTTING POSITIONS
*
*  IMSL SUBROUTINES:
*       GGNLG - GENERATES LOGNORMAL RANDOM DEVIATES
```

```
*       VRSTA - ORDERS DATA IN ASCENDING VALUE
*       MDNOR - CALCULATES THE NORMAL PDF OF AN OBSERVATION
*
*  OUTPUTS:
*       KSCRIT = 3-D ARRAY OF CRITICAL VALUES FOR MOD. K-S TEST
*       ADCRIT = 3-D ARRAY OF CRITICAL VALUES FOR MOD. A-D TEST
*       CVCRIT = 3-D ARRAY OF CRITICAL VALUES FOR MOD.C-VM TEST
*
***************************************************************
***************************************************************
C
        PROGRAM CRITICAL
C
        COMMON DSEED,X,N,C,NC,AMLE,BMLE,P,PCT,
     1       KS,AD,CVM,IT,NSIZ,NSHP,NPCT,NST,
     1       KSCRIT,ADCRIT,CVCRIT,Y
        INTEGER N,NSIZ,NSHP,IT,NPCT,NST
        REAL X(31),AMLE,BMLE,KS(5000,6,7),AD(5000,6,7),
     1       CVM(5000,6,7),C,NC,P(30),
     1       KSCRIT(6,8,5),ADCRIT(6,8,5),CVCRIT(6,8,5),PCT,
     1       Y(5002),ALPHA
       DOUBLE PRECISION DSEED
C
**  OPEN OUTPUT FILES TO STORE COMPUTED CRITICAL VALUES:
C
       OPEN (UNIT=7,FILE='CRITICAL',STATUS='NEW')
C
**  NUMBER OF TEST STATISTICS TO BE USED ON EACH RUN:
C
C       PRINT*,'THE MONTE CARLO ANALYSIS WILL REQUIRE'
C       PRINT*,'      5000 TEST STATISTICS.'
C       PRINT*,'ENTER THE NUMBER TO BE USED FOR THIS RUN:'
C       READ*,NST
C
      NST = 5000
C
      CALL FILL
C
**  CALCULATE 5002 PLOTTING POSITIONS ON THE Y-AXIS:
C
       Y(0) = 0.0
       DO 10 I = 1,NST
          Y(I) = (I-0.3)/(NST + 0.4)
 10    CONTINUE
C
       Y(NST + 1) = 1.0
C
C      PRINT*,'ENTER RANDOM NUMBER SEED OR "1." FOR DEFAULT:'
C      READ*,DSEED
C      IF (DSEED .EQ. 1.) DSEED = 123457.0D0
C      PRINT*,'  '
C      PRINT*,'STANDBY . . . COMPUTATIONS IN PROGRESS'
C
```

```
      DSEED = 123457.0D0
C
      NSHP = 0
C
** BEGIN DO LOOP 90 FOR SHAPE PARAMETER VALUES C=1.0(.5)4.0  **
C
      DO 90 SHAPE = 1.0,4.0,.5
         C = SHAPE
         NSHP = NSHP + 1
C
**   WRITE HEADINGS FOR OUTPUT DATA:
C
         WRITE(7,52)
         WRITE(7,51)
         WRITE(7,52)
         WRITE(7,54)
         WRITE(7,52)
         WRITE(7,56)
C
         NSIZ = 0
C
**  BEGIN DO LOOP 80 FOR SAMPLE SIZES N=5(5)30
C
         DO 80 NSAMP = 5,30,5
            N = NSAMP
            NSIZ = NSIZ + 1
            NC = N * C
C
            WRITE(7,58)
C
***      BEGIN DO LOOP 60 FOR 5000 ITERATIONS   ***
C
            DO 60 IT = 1,NST
C
**             PERFORM STEPS 1&2 OF FIG 6:   **
C
               CALL LOGDEV
C
**             PERFORM STEP 3 OF FIGURE 6:   **
C
               CALL MLE
C
**             PERFORM STEP 4 OF FIGURE ***
C
               CALL HYPCDF
C
**             PERFORM STEP 5 OF FIGURE ***
C
               CALL TESTAT
C
 60         CONTINUE
C
**     END DO LOOP 60 FOR 5000 ITERATIONS  **
```

```
C
**     PERFORM STEP 7 OF FIGURE 6:         **
C
             DO 70 NPCT = 1,5
C
                CALL CRTVAL
C
                WRITE(7,62),1.-PCT,N,C,KSCRIT(NSIZ,NSHP,NPCT),
     1            ADCRIT(NSIZ,NSHP,NPCT),CVCRIT(NSIZ,NSHP,NPCT)
C
 70          CONTINUE
C
***          END DO LOOP 70 FOR PERCENTILES    ***
C
 80       CONTINUE
C
***       END DO LOOP 80 FOR PERCENTILES    ***
C
 90    CONTINUE
C
***  END DO LOOP 90 FOR SHAPE PARAMETER VALUES C=1.0(.5)4.0 ***
C
***************************************************************
*** OUTPUT INSTRUCTIONS:   THE FOLLOWING FORMATS THE OUTPUT ***
***     THE DATA TO A FILE TO BE PRINTED OUT IN HARDCOPY    ***
***************************************************************
C
***  WRITE KS CRITICAL VALUE TABLES TO FILE BY ALPHA LEVEL  ***
C
       WRITE(7,52)
       WRITE(7,130)
       WRITE(7,52)
       WRITE(7,132)
       WRITE(7,52)
       WRITE(7,200)
       WRITE(7,201)
       WRITE(7,52)
C
       NPCT = 0
C
*** BEGIN DO LOOP 105 TO SORT CRITICAL VALUES BY ALPHA LEVEL***
C
       DO 105 NPCT = 1,5
C
          IF (NPCT .NE. 5) ALPHA = .25 - (.05*NPCT)
          IF (NPCT .EQ. 5) ALPHA = .01
C
          NSIZ = 0
             N = 0
C
***       BEGIN DO LOOP 107 TO SORT OUTPUT BY SAMPLE SIZE  ***
C
          DO 107 NSIZ = 1,6
```

```
C
                N = 5 * NSIZ
C
                WRITE(7,120),ALPHA,N,KSCRIT(NSIZ,1,NPCT),KSCRIT
     1           (NSIZ,2,NPCT),KSCRIT(NSIZ,3,NPCT),KSCRIT(NSIZ,
     1           4,NPCT),KSCRIT(NSIZ,5,NPCT),KSCRIT(NSIZ,6,NPCT),
     1           KSCRIT(NSIZ,7,NPCT)
C
 107         CONTINUE
C
***   END DO LOOP 107 AFTER SORTING OUTPUT BY SAMPLE SIZE   ***
C
             WRITE(7,201)
C
 105    CONTINUE
C
***   END DO LOOP 105 AFTER SORTING OUTPUT BY ALPHA LEVEL   ***
C
***   WRITE AD CRITICAL VALUE TABLES TO FILE BY ALPHA LEVEL ***
C
        WRITE(7,52)
        WRITE(7,140)
        WRITE(7,52)
        WRITE(7,142)
        WRITE(7,52)
        WRITE(7,200)
        WRITE(7,201)
        WRITE(7,52)
C
        NPCT = 0
C
*** BEGIN DO LOOP 115 TO SORT CRITICAL VALUES BY ALPHA LEVEL***
C
        DO 115 NPCT = 1,5
C
        IF (NPCT .NE. 5) ALPHA = .25 - (.05*NPCT)
        IF (NPCT .EQ. 5) ALPHA = .01
C
        NSIZ = 0
        N = 0
C
***     BEGIN DO LOOP 117 TO SORT OUTPUT BY SAMPLE SIZE  ***
C
        DO 117 NSIZ = 1,6
        N = 5 * NSIZ
C
        WRITE(7,120),ALPHA,N,ADCRIT(NSIZ,1,NPCT),ADCRIT
     1      (NSIZ,2,NPCT),ADCRIT(NSIZ,3,NPCT),ADCRIT(NSIZ,
     1      4,NPCT),ADCRIT(NSIZ,5,NPCT),ADCRIT(NSIZ,6,NPCT),
     1      ADCRIT(NSIZ,7,NPCT)
C
 117    CONTINUE
C
***     END DO LOOP 117 AFTER SORTING BY SAMPLE SIZE  ***
```

```
C
      WRITE(7,201)
C
 115  CONTINUE
C
***   END DO LOOP 115 AFTER SORTING OUTPUT BY ALPHA LEVEL  ***

C
      WRITE(7,52)
      WRITE(7,150)
      WRITE(7,52)
      WRITE(7,152)
      WRITE(7,52)
      WRITE(7,200)
      WRITE(7,201)
      WRITE(7,52)
C
      NPCT = 0
C
*** BEGIN DO LOOP 125 TO SORT CRITICAL VALUES BY ALPHA LEVEL***
C
      DO 125 NPCT = 1,5
C
      IF (NPCT .NE. 5) ALPHA = .25 - (.05*NPCT)
      IF (NPCT .EQ. 5) ALPHA = .01
C
      NSIZ = 0
      N = 0
C
*** BEGIN DO LOOP 127 TO SORT CRITICAL VALUES BY ALPHA LEVEL***
C
      DO 127 NSIZ = 1,6
      N = 5 * NSIZ
C
      WRITE(7,120),ALPHA,N,CVCRIT(NSIZ,1,NPCT),CVCRIT
     1     (NSIZ,2,NPCT),CVCRIT(NSIZ,3,NPCT),CVCRIT(NSIZ,
     1     4,NPCT),CVCRIT(NSIZ,5,NPCT),CVCRIT(NSIZ,6,NPCT),
     1     CVCRIT(NSIZ,7,NPCT)
C
 127  CONTINUE
C
***   END DO LOOP 127 AFTER SORTING BY SAMPLE SIZE   ***
C
      WRITE(7,201)
C
 125  CONTINUE
C
***   END DO LOOP 125 AFTER SORTING OUTPUT BY ALPHA LEVEL  ***
C
```

```
  51    FORMAT(' ********************************************')
  52    FORMAT('  ')
  54    FORMAT(' ''LOGNORMAL CRITICAL VALUES FOR SHAPE C =  **')
  56    FORMAT(' ALPHA',3X,'N',4X,'C',7X,'KS',8X,'AD',8X,'CVM')
  58    FORMAT('----------------------------------------------')
  62    FORMAT(' ',T3,F3.2,I5,F6.1,3F10.4)
 120    FORMAT(' ',T3,F3.2,I5,F8.3,7F9.3)
 130    FORMAT('1',36X,'TABLE I')
 132    FORMAT(20X,'CRITICAL VALUES FOR THE MODIFIED K-S TEST')
 140    FORMAT('1',36X,'TABLE II')
 142    FORMAT(20X,'CRITICAL VALUES FOR THE MODIFIEL A-D TEST')
 150    FORMAT('1',35X,'TABLE III')
 152    FORMAT(19X,'CRITICAL VALUES FOR THE MODIFIED C-VM TEST')
 200    FORMAT(' ALPHA',3X,'N',4X,'C=1.0',5X,'1.5',6X,
     1             '2.0',6X,'2.5',6X,'3.0',6X,'3.5',6X,'4.0')
 201    FORMAT(73('-'))
C
        CLOSE(7)
C
        END
C
**        END MAIN PROGRAM      ****
C
```

```
*****************************************************************
*
*  PURPOSE:  TO FILL ALL ARRAYS USED IN THIS PROGRAM WITH THE
*            VALUE OF 0
*
*  VARIABLES:
*                  X = ARRAY OF LOGNORMAL RANDOM DEVIATES
*                  P = ARRAY OF N POINTS OF HYPOTHESIZED CDF
*                 KS = ARRAY OF VALUES OF MOD. K-S TEST STATISTIC
*                 AD = ARRAY OF VALUES OF MOD. A-D TEST STATISTIC
*                CVM = ARRAY OF VALUES OF MOD. C-VM TEST STATISTIC
*             KSCRIT = ARRAY OF K-S CRITICAL VALUES
*             ADCRIT = ARRAY OF A-D CRITICAL VALUES
*             CVCRIT = ARRAY OF C-VM CRITICAL VALUES
*
*****************************************************************
C
      SUBROUTINE FILL
C
       COMMON DSEED,X,N,C,NC,AMLE,BMLE,P,PCT,
     1       KS,AD,CVM,IT,NSIZ,NSHP,NPCT,NST,
     1       KSCRIT,ADCRIT,CVCRIT,Y
       INTEGER N,NSIZ,NSHP,IT,NPCT,NST
       REAL X(31),AMLE,BMLE,KS(5000,6,7),AD(5000,6,7),
     1       CVM(5000,6,7),C,NC,P(30),
     1       KSCRIT(6,8,5),ADCRIT(6,8,5),CVCRIT(6,8,5),PCT,
     1       Y(5002),ALPHA
       DOUBLE PRECISION DSEED
C
      DO 10 I=1,31
         X(I) = 0.0
 10   CONTINUE
C
      DO 20 I=1,30
         P(I) = 0.0
 20   CONTINUE
C
      DO 30 I=1,6
C
         DO 40 J=1,7
C
            DO 50 K=1,5000
C
               KS(K,I,J)=0.0
               AD(K,I,J)=0.0
              CVM(K.I,J)=0.0
C
 50         CONTINUE
C
            DO 60 L=1,5
C
                 KSCRIT(I,J,L)=0.0
                 ADCRIT(I,J,L)=0.0
```

```
                  CVCRIT(I,J,L)=0.0
C
 60             CONTINUE
C
 40          CONTINUE
C
 30      CONTINUE
C
         RETURN
C
         END
C
***      END SUBROUTINE FILL    ***
C
```

```
*****************************************************************
*
*  PURPOSE:  TO GENERATE N RANDOM DEVIATES FROM A LOGNORMAL
*            DISTRIBUTION WHOSE PARENT NORMAL HAS MEAN OF 0 AND
*            STANDARD DEVIATION OF 1.  THE PROGRAM THEN ADDS A
*            LOCATION PARAMETER OF 10 TO EACH DEVIATE TO
*            PRODUCE THE THREE-PARAMETER LOGNORMAL
*            DEVIATE FROM THE TWO-PARAMETER LOGNORMAL.
*
*  VARIABLES:
*               DSEED = RANDOM NUMBER SEED
*                   X = ARRAY OF LOGNORMAL RANDOM DEVIATES
*                   N = SAMPLE SIZE
*                 PMU = MEAN OF PARENT NORMAL OF LOGNORMAL
*                PVAR = VARIANCE OF PARENT NORMAL OF LOGNORMAL
*
*  IMSL SUBROUTINES:
*               GGLNG - GENERATES LOGNORMAL RANDOM DEVIATES
*               VRSTA - ORDERS DATA IN ASCENDING VALUE
*
*****************************************************************
C
      SUBROUTINE LOGDEV
C
      COMMON DSEED,X,N,C,NC,AMLE,BMLE,P,PCT,
     1      KS,AD,CVM,IT,NSIZ,NSHP,NPCT,NST,
     1      KSCRIT,ADCRIT,CVCRIT,Y
      INTEGER N,NPCT,NSIZ,NSHP,IT,NST
      REAL X(31),AMLE,BMLE,KS(5000,6,7),AD(5000,6,7),
     1      CVM(5000,6,7),C,NC,P(30),
     1      KSCRIT(6,8,5),ADCRIT(6,8,5),CVCRIT(6,8,5),
     1      Y(5002),PCT,ALPHA
     DOUBLE PRECISION DSEED
C
     REAL PMU,PVAR
C
      PMU  = 0.0
      PVAR = 1.0
C
      CALL GGNLG(DSEED,N,PMU,PVAR,X)
C
***  ADD THE LOCATION PARAMETER OF 10 TO DEVIATES   ***
C
      DO 10 I=1,N
         X(I) = X(I) + 10.0
 10   CONTINUE
C
      CALL VSRTA(X,N)
C
      RETURN
C
      END
C
***  END SUBROUTINE LOGDEV   ***
```

```
C
*****************************************************************
*
*  PURPOSE:  TO ESTIMATE THE LOCATION AND THE SCALE PARAMETERS
*            FROM THE SAMPLE DATA USING A BI-SECTION SEARCH.
*
*  VARIABLES:
*                X = ARRAY OF LOGNORMAL RANDOM DEVIATES
*                N = SAMPLE SIZE
*             AMLE = MLE OF THE LOCATION PARAMETER
*             BMLE = MLE OF THE SCALE PARAMETER
*              DIF = VARIABLE USED IN CALCULATIONS
*             TDIF = VARIABLE USED IN CALCULATIONS
*             TEMP = VARIABLE USED IN CALCULATIONS
*               UP = UPPER BOUND OF LOCATION PARAMETER
*            UPPER = VALUE RETURNED BY CALC FOR UP
*              LOW = LOWER OF STEPS IN BISECTION SEARCH
*            LOWER = VALUE RETURNED BY CALC FOR LOW
*              MID = VALUE OF MID-POINT BETWEEN UP AND LOW
*           MIDDLE = VALUE RETURNED BY CALC FOR MID
*             STEP = SIZE OF BACKWARD STEP = 10% OF X(1)
*            THETA = VARIABLE USED IN CALCULATIONS
*           STHETA = SUM OF ALL THETA
*           HTHETA = LOGNORMAL OF THE  ESTIMATE FOR THE SCALE PAR.
*
*  SUBROUTINES:
*             CALC - PERFORMS NECESSARY CALCULATIONS FOR MLE
*
*****************************************************************
C
      SUBROUTINE MLE
C
       COMMON DSEED,X,N,C,NC,AMLE,BMLE,P,PCT,
     1       KS,AD,CVM,IT,NSIZ,NSHP,NPCT,NST,
     1       KSCRIT,ADCRIT,CVCRIT,Y
       INTEGER N,NSIZ,NSHP,IT,NPCT,NST
       REAL X(31),AMLE,BMLE,KS(5000,6,7),AD(5000,6,7),
     1       CVM(5000,6,7),C,NC,P(30),
     1       KSCRIT(6,8,5),ADCRIT(6,8,5),CVCRIT(6,8,5),
     1       Y(5002),PCT,ALPHA
      DOUBLE PRECISION DSEED
C
      REAL LOW,LOWER,MID,MIDDLE,UP,UPPER,TEMP,STEP,THETA,
     1         STHETA,HTHETA
C
      DIF = 0.0
      TDIF = 0.0
C
      UP = X(1)
C
      CALL CALC(UP,X,N,C,UPPER)
C
      STEP = (.1*X(1))
```

```
C
      LOW = X(1)-STEP
C
 5    CONTINUE
C
      CALL CALC(LOW,X,N,C,LOWER)
C
      IF ((UPPER*LOWER) .GT. 0.0) THEN
         UP = LOW
         LOW = LOW-STEP
         UPPER = LOWER
         GO TO 5
      END IF
C
 10   CONTINUE
C
      MID = (UP+LOW)/2
C
      CALL CALC(MID,X,N,C,MIDDLE)
C
      IF ((UPPER * MIDDLE) .LE. 0.0) THEN
         LOW = MID
         LOWER = MIDDLE
      ELSE
         UP = MID
         UPPER = MIDDLE
      END IF
C
      IF (ABS(UP-LOW) .GT. .01) GO TO 10
C
      AMLE = MID
C
      STHETA = 0.0
C
      DO 15 I=1,N
         TEMP = LOG(X(I) - AMLE)
         STHETA = STHETA + TEMP
 15   CONTINUE
C
      HTHETA = STHETA/N
      BMLE = HTHETA
C
      RETURN
C
      END
C
***     END SUBROUTINE MLE   ***
C
```

```
******************************************************************
*
*  PURPOSE:  TO PREFORM THE NECESSARY CALCULATIONS FOR THE MLE
*            SEARCH.
*
*  VARIABLES:
*       DSEED = RANDOM NUMBER SEED
*         LOC = CURRENT LOCATION PARAMETER USED IN CALCULATIONS
*           X = ARRAY OF LOGNORMAL RANDOM VARIABLES
*           N = SAMPLE SIZE
*         SHP = CURRENT SHAPE PARAMETER USED IN CALCULATIONS
*        TSUM = VARIABLE USED IN CALCULATIONS
*         DIF = VARIABLE USED IN CALCULATIONS
*        TDIF = VARIABLE USED IN CALCULATIONS
*         SUM = VARIABLE USED IN CALCULATIONS
*       LNDIF = VARIABLE USED IN CALCULATIONS
*
******************************************************************
C
      SUBROUTINE CALC(LOC,X,N,SHP,TSUM)
C
      INTEGER N
      REAL LOC,X(31),SHP,TSUM,DIF,SUM,TDIF,LNDIF
C
      DOUBLE PRECISION DSEED
C
      SUM = 0.0
      TSUM = 0.0
C
      DO 5 I=1,N
         DIF = X(I)-LOC
         IF (DIF .EQ. 0.0) DIF = .00001
         LNDIF = LOG(DIF)
         SUM = SUM + LNDIF
 5    CONTINUE
C
      SUM = SUM/N
C
      DO 10 I=1,N
         DIF = X(I) - LOC
         IF (DIF .EQ. 0.0) DIF = .00001
         TDIF = 1/DIF
         LNDIF = LOG(DIF)
         TSUM = TSUM+TDIF+(1/SHP)*TDIF*(LNDIF-SUM)
 10   CONTINUE
C
      RETURN
C
      END
C
***  END SUBROUTINE CALC   ***
C
```

```
*******************************************************************
*
*  PURPOSE:  GIVEN AN ORDERED SAMPLE OF SIZE N, A SPECIFIED
*            SHAPE C, AND THE MLE OF THE LOCATION AND SCALE,
*            COMPUTE THE HYPOTHESIZED LOGNORMAL DISTRIBUTION
*            FUNCTION L(I) FOR I = 1,2,...N.
*
*  VARIABLES:
*               C = SHAPE PARAMETER
*               X = ARRAY OF LOGNORMAL RANDOM DEVIATES
*               N = SAMPLE SIZE
*            AMLE = MLE OF THE LOCATION PARAMETER
*            BMLE = MLE OF THE SCALE PARAMETER
*               P = ARRAY OF N POINTS OF HYPOTHESIZED CDF
*
*  IMSL SUBROUTINE:
*        MDNOR - CALCULATES THE NORMAL PDF OF AN OBSERVATION
*
*******************************************************************
C
       SUBROUTINE HYPCDF
C
       COMMON DSEED,X,N,C,NC,AMLE,BMLE,P,PCT,
     1       KS,AD,CVM,IT,NSIZ,NSHP,NPCT,NST,
     1       KSCRIT,ADCRIT,CVCRIT,Y
       INTEGER N,NSIZ,NSHP,IT,NPCT,NST
       REAL X(31),AMLE,BMLE,KS(5000,6,7),AD(5000,6,7),
     1       CVM(5000,6,7),C,NC,P(30),
     1       KSCRIT(6,8,5),ADCRIT(6,8,5),CVCRIT(6,8,5),
     1       Y(5002),ALPHA,PCT
      DOUBLE PRECISION DSEED
C
      REAL Q,Z
C
       DO 10 I = 1,N
           Q = (LOG(X(I)-AMLE)-BMLE)/C
           CALL MDNOR(Q,Z)
           P(I) = Z
 10    CONTINUE
C
       RETURN
C
       END
C
***   END SOUROUTINE HYPCDF  ***
C
```

```
***************************************************************
*
*  PURPOSE:  GIVEN A SAMPLE SIZE N, AND THE HYPOTHESIZED
*            LOGNORMAL DESTRIBUTION FUNCTION L(I), COMPUTE
*            VALUES OF THE TEST STATISTICS OF THE MODIFIED K-S,
*            A-D, AND CVM GOODNESS-OF-FIT TESTS.
*
*  VARIABLES:
*       N = SAMPLE SIZE
*    NSHP = SHAPE PARAMETER COUNTER
*    NSIZ = SAMPLE SIZE COUNTER
*      IT = ITERATION COUNTER (1-5000)
*       P = ARRAY OF N VALUES OF THE HYPOTHESIZED LOGNORMAL CDF
*
*      DP = POSITIVE DIFFERENCES BETWEEN EDF AND CDF POINTS
*      DM = NEGATIVE DIFFERENCES BETWEEN EDF AND CDF POINTS
*   DPLUS = MAXIMUM POSITIVE DIFFERENCE (LARGEST DP VALUE)
*  DMINUS = MAXIMUM NEGATIVE DIFFERENCE (LARGEST DM VALUE)
*      KS = VALUES OF THE MODIFIED K-S TEST STATISTIC
*
*      AL = VALUE USED TO CALCULATE THE A-D TEST STATISTIC
*      AM = VALUE USED TO CALCULATE THE A-D TEST STATISTIC
*      AN = AL + AM
*     AAA = VALUES TO BE SUMMED FOR A-D TEST STATISTIC
*    SAAA = SUM OF AAA VALUES
*      AD = VALUES OF THE MODIFIED A-D TEST STATISTIC
*
*     ACV = SQUARED QUANTITIES IN THE C-VM FORMULA
*    SACV = SUM OF THE ACV VALUES
*     CVM = VALUES OF THE MODIFIED C=VM TEST STATISTIC
*
***************************************************************
C
      SUBROUTINE TESTAT
C
      COMMON DSEED,X,N,C,NC,AMLE,BMLE,P,PCT,
     1       KS,AD,CVM,IT,NSIZ,NSHP,NPCT,NST,
     1       KSCRIT,ADCRIT,CVCRIT,Y
      INTEGER N,NSIZ,NSHP,IT,IK,NPCT,NST
      REAL X(31),AMLE,BMLE,KS(5000,6,7),AD(5000,6,7),
     1       CVM(5000,6,7),C,NC,P(30),
     1       KSCRIT(6,8,5),ADCRIT(6,8,5),CVCRIT(6,8,5),
     1       DP(30),DM(30),DPLUS,DMINUS,AL(30),AM(30),PCT,
     1       AN(30),AAA(30),SAAA,ACV(30),SACV,Y(5002),ALPHA
     DOUBLE PRECISION DSEED
C
      DPLUS = 0
      DMINUS = 0
C
      DO 5 IK = 1,30
         DP(IK) = 0
         DM(IK) = 0
 5    CONTINUE
```

```
***    COMPUTE THE K-S TEST STATISTIC  ***
C
       DO 10 I = 1,N
          DP(I) = ABS( (I/REAL(N)) - P(I) )
          DM(I) = ABS( P(I) - (I-1)/REAL(N) )
 10    CONTINUE
C
       DPLUS = MAX( DP(1),DP(2),DP(3),DP(4),DP(5),DP(6),DP(7),
     1       DP(8),DP(9),DP(10),DP(11),DP(12),DP(13),DP(14),
     1       DP(15),DP(16),DP(17),DP(18),DP(19),DP(20),
     1       DP(21),DP(22),DP(23),DP(24),DP(25),DP(26),
     1       DP(27),DP(28),DP(29),DP(30) )
C
       DMINUS = MAX( DM(1),DM(2),DM(3),DM(4),DM(5),DM(6),DM(7),
     1       DM(8),DM(9),DM(10),DM(11),DM(12),DM(13),DM(14),
     1       DM(15),DM(16),DM(17),DM(18),DM(19),DM(20),
     1       DM(21),DM(22),DM(23),DM(24),DM(25),DM(26),
     1       DM(27),DM(28),DM(29),DM(30) )
C
       KS(IT,NSIZ,NSHP) = MAX(DPLUS,DMINUS)
C
***    COMPUTE THE A-D TEST STATISTIC    ***
C
       SAAA = 0
C
       DO 20 J = 1,N
          IF (P(J) .LE. .001) P(J) = .001
          AL(J) = LOG (P(J))
          IF (P(N+1-J) .LE. .001) P(N+1-J) = .001
          AM(J) = LOG (1.0 - P(N+1-J))
          AN(J) = AL(J) + AM(J)
          AAA(J) = (2.0*J - 1.0) * AN(J)
          SAAA = SAAA + AAA(J)
 20    CONTINUE
C
       AD(IT,NSIZ,NSHP) = -N - (1.0/REAL(N)) * SAAA
C
***    COMPUTE THE C-VM TEST STATISTIC   ***
C
       SACV = 0
C
       DO 30 K = 1,N
          ACV(K) = ( P(K) - (2.0*K - 1.0)/(2.0*REAL(N)) )**2
          SACV = SACV + ACV(K)
 30    CONTINUE
C
       CVM(IT,NSIZ,NSHP) = SACV + (1.0/(12.0*REAL(N)))
C
       RETURN
C
       END
C
***  END SUBROUTINE TESTAT  ***
C
```

```
****************************************************************
*
*  PURPOSE:  CALCULATES THE CRITIAL VALUES FOR A GIVEN LEVEL OF
*            SIGNIFICANCE
*
*  VARIABLES:
*           C = SHAPE PARAMETER
*           N = SAMPLE SIZE
*        NSHP = SHAPE PARAMETER COUNTER
*        NSIZ = SAMPLE SIZE COUNTER
*        NPCT = PERCENTILE COUNTER
*         NST = TOTAL NUMBER OF STATISTICS USED
*          IT = ITERATION COUNTER (5000 REQUIRED)
*          KS = ARRAY OF VALUES OF MODIFIED K-S TEST STATISTIC
*         CVM = ARRAY OF VALUES OF MODIFIED C-VM TEST STATISTIC
*          AD = ARRAY OF VALUES OF MODIFIED A-D TEST STATISTIC
*       ALPHA = LEVEL OF SIGNIFICANCE
*
*          KS = 3-D ARRAY OF 5000 MODIFIED K-S TEST STATISTICS
*         KS1 = 1-D ARRAY OF 5000 K-S TEST STATISTICS
*         CVM = 3-D ARRAY OF 5000 MODIFIED C-VM TEST STATISTICS
*         CV1 = 1-D ARRAY OF 5000 C-VM TEST STATISTICS
*          AD = 3-D ARRAY OF 5000 MODIFIED A-D TEST STATISTICS
*         AD1 = 1-D ARRAY OF 5000 A-D TEST STATISTICS
*        STAT = 1-D ARRAY OF TEST STATS (KS, AD, OR CVM)
*      KSCRIT = ARRAY OF CRITICAL VALUES FOR THE K-S TEST
*     CVMCRIT = ARRAY OF CRITICAL VALUES FOR THE C-VM TEST
*      ADCRIT = ARRAY OF CRITICAL VALUES FOR THE A-D TEST
*        CRIT = EITHER THE KS, AD, OR CVM CRITICAL VALUE ARRAY
*           Y = ARRAY CONTAINING 5002 PLOTTING POSITIONS
*        SLPM = ARRAY OF SLOPES USED TO FIND CRITICAL VALUES
*          BI = ARRAY OF INTERCEPTS USED TO FIND CRITICAL VALS.
*
*  SUBROUTINE:
*          VRSTA - ORDERS DATA IN ASCENDING VALUE
*
****************************************************************
C
      SUBROUTINE CRTVAL
C
      COMMON DSEED,X,N,C,NC,AMLE,BMLE,P,PCT,
     1      KS,AD,CVM,IT,NSIZ,NSHP,NPCT,NST,
     1      KSCRIT,ADCRIT,CVCRIT,Y
      INTEGER N,NSIZ,NSHP,IT,NPCT,NST,NTEST
      REAL X(31),AMLE,BMLE,KS(5000,6,7),AD(5000,6,7),
     1      CVM(5000,6,7),C,NC,P(30),
     1      KSCRIT(6,8,5),ADCRIT(6,8,5),CVCRIT(6,8,5),PCT,
     1      Y(5002),STAT(5002),CRIT(6,8,7),SLPM(7),BI(7),
     1      KS1(5000),CV1(5000),AD1(5000),ALPHA
     DOUBLE PRECISION DSEED
C
      IF (NPCT .EQ. 1) PCT = .80
      IF (NPCT .EQ. 2) PCT = .85
```

```
      IF (NPCT .EQ. 3) PCT = .90
      IF (NPCT .EQ. 4) PCT = .95
      IF (NPCT .EQ. 5) PCT = .99
C
***  STORE THE 3 SETS OF 5000 TEST STATS INTO 1-D ARRAYS:   ***
C
      DO 16 NCNT = 1,NST
         KS1(NCNT) = KS(NCNT,NSIZ,NSHP)
         AD1(NCNT) = AD(NCNT,NSIZ,NSHP)
         CV1(NCNT) = CVM(NCNT,NSIZ,NSHP)
 16   CONTINUE
C
***   USE IMSL SUBROUTINE TO ORDER THE TEST STATISTICS:   ***
C
      CALL VSRTA(KS1,NST)
C
      CALL VSRTA(AD1,NST)
C
      CALL VSRTA(CV1,NST)
C
***   BEGIN DO LOOP 20 TO ROTATE THROUGH KS, AD, CVM   ***
C
      DO 20 NTEST = 1,3
C
***   BEGIN DO LOOP 30 FOR 5000 DATA POINTS   ***
C
      DO 30 J = 1,NST
         IF (NTEST .EQ. 1) THEN
             STAT(J) = KS1(J)
         ELSE IF (NTEST .EQ. 2) THEN
             STAT(J) = AD1(J)
         ELSE IF (NTEST .EQ. 3) THEN
             STAT(J) = CV1(J)
         END IF
 30   CONTINUE
C
***   END DO LOOP 30 FOR 5000 DATA POINTS   ***
C
***   EXTRAPOLATE LEFT ENDPOINT OF THE TEST STATISTICS:   ***
C
      IF (STAT(1) .EQ.  STAT(2)) THEN
        DIFO = STAT(3) - STAT(1)
        IF (DIFO .EQ. 0.0) DIFO = .00001
        SLPM(0) = (Y(3) - Y(1))/DIFO
      ELSE
        DIFO = STAT(2) - STAT(1)
        SLPM(0) = (Y(2) - Y(1)) / DIFO
      END IF
C
      BI(0) = Y(1) - SLPM(0) * STAT(1)
      STAT(0) = MAX(0.0, - BI(0)/SLPM(0) )
C
***   EXTRAPOLATE RIGHT ENDPOINT OF THE TEST STATISTIC   ***
```

```
C
        IF (STAT(NST-1) .EQ. STAT(NST)) THEN
           DIF6 = STAT(NST) - STAT(NST-2)
           IF (DIF6 .EQ.0.0) DIF6 = .00001
           SLPM(6) = (Y(NST)-Y(NST-2)) / DIF6
        ELSE
           DIF6 = STAT(NST) - STAT(NST-1)
           SLPM(6) = (Y(NST)-Y(NST-1)) / DIF6
        END IF
C
        BI(6) = Y(NST-1) - SLPM(6)*STAT(NST-1)
        STAT(NST+1) = (1.0 - BI(6)) / SLPM(6)
C
***     INTERPOLATE CRITICAL VALUES BETWEEN TEST STATS:     ***
C
***     BEGIN DO LOOP 50 TO FIND MAX Y(K) < PCT:            ***
C
        DO 50 KJ = 1,NST
           K = NST+1 - KJ
C
           IF (Y(K) .LE. PCT) THEN
C
              IF (STAT(K) .EQ. STAT(K+1)) THEN
                 DIF = STAT(K+1) - STAT(K-1)
                 IF (DIF .EQ.0.0) DIF = .00001
                 SLPM(NPCT) = (Y(K+1)-Y(K-1)) / DIF
              ELSE
                 DIF = STAT(K+1) - STAT(K)
                 SLPM(NPCT) = (Y(K+1)-Y(K)) / DIF
              END IF
C
              BI(NPCT) = Y(K) - SLPM(NPCT) * STAT(K)
              CRIT(NSIZ,NSHP,NPCT)
     1                 = (PCT-BI(NPCT))/SLPM(NPCT)
C
              GO TO 75
           END IF
C
 50     CONTINUE
C
***     END DO LOOP 50 UPON FINDING CRIT VAL     ***
C
***     ASSOCIATE THE CRITICAL VALUES WITH TEST TYPES    ***
C
 75     IF (NTEST .EQ. 1) THEN
           KSCRIT(NSIZ,NSHP,NPCT) = CRIT(NSIZ,NSHP,NPCT)
        ELSE IF (NTEST .EQ. 2) THEN
           ADCRIT(NSIZ,NSHP,NPCT) = CRIT(NSIZ,NSHP,NPCT)
        ELSE IF (NTEST .EQ.3) THEN
           CVCRIT(NSIZ,NSHP,NPCT) = CRIT(NSIZ,NSHP,NPCT)
        END IF
C
 20     CONTINUE
```

```
C
***  END DO LOOP 20 AFTER ROTATING THROUGH KS, AD, AND CVM  ***
C
      RETURN
C
      END
C
***   END SUBROUTINE CRTVALL   ***
```

APPENDIX B

Flow Chart for Program Power

Computer Program and Subroutines
for Determining Power
Values

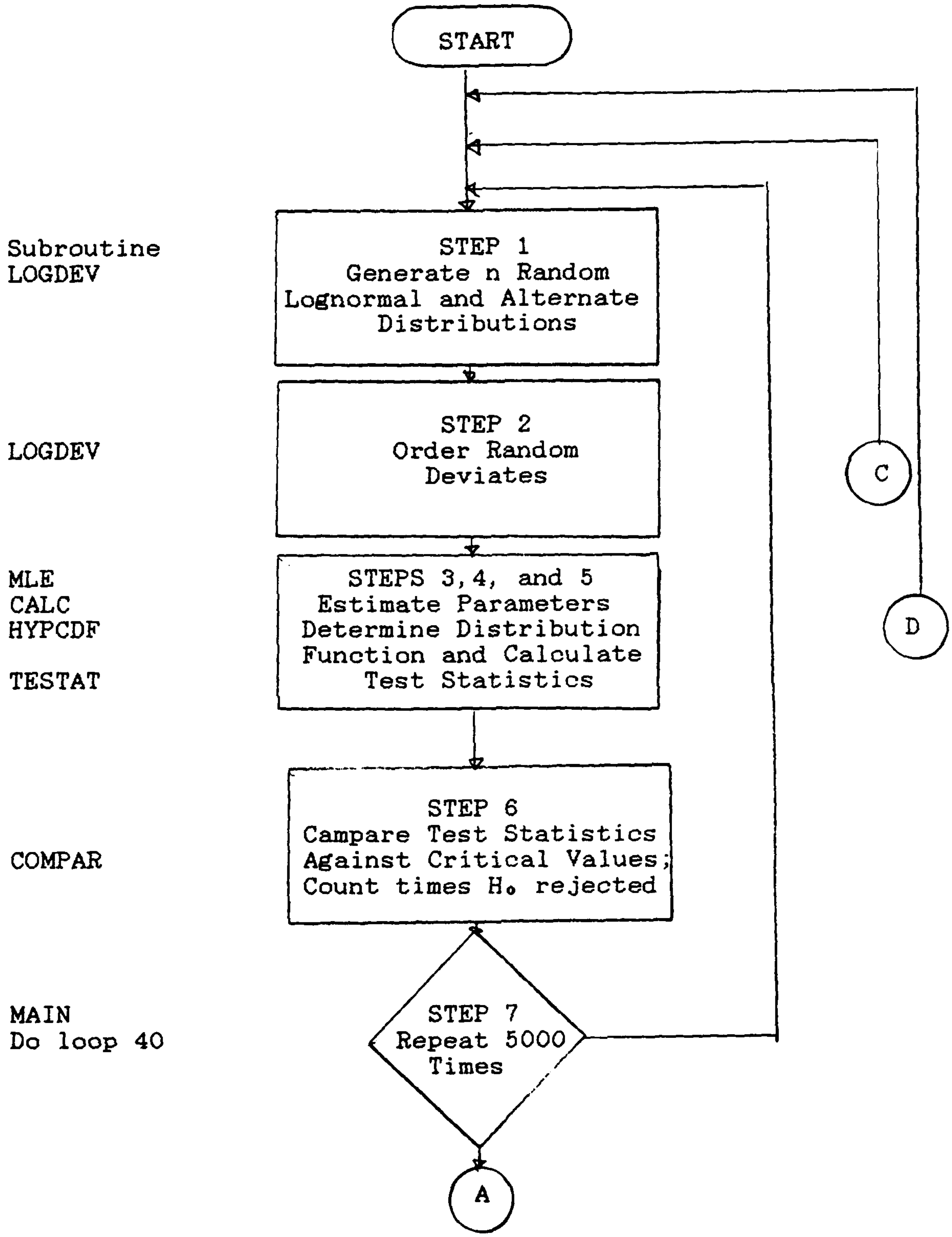


Fig 2. Flow chart for Program POWER

MAIN

MAIN
Do loop 60

MAIN
Do loop 7

MAIN
Do loop 80

MAIN
Do loop 90

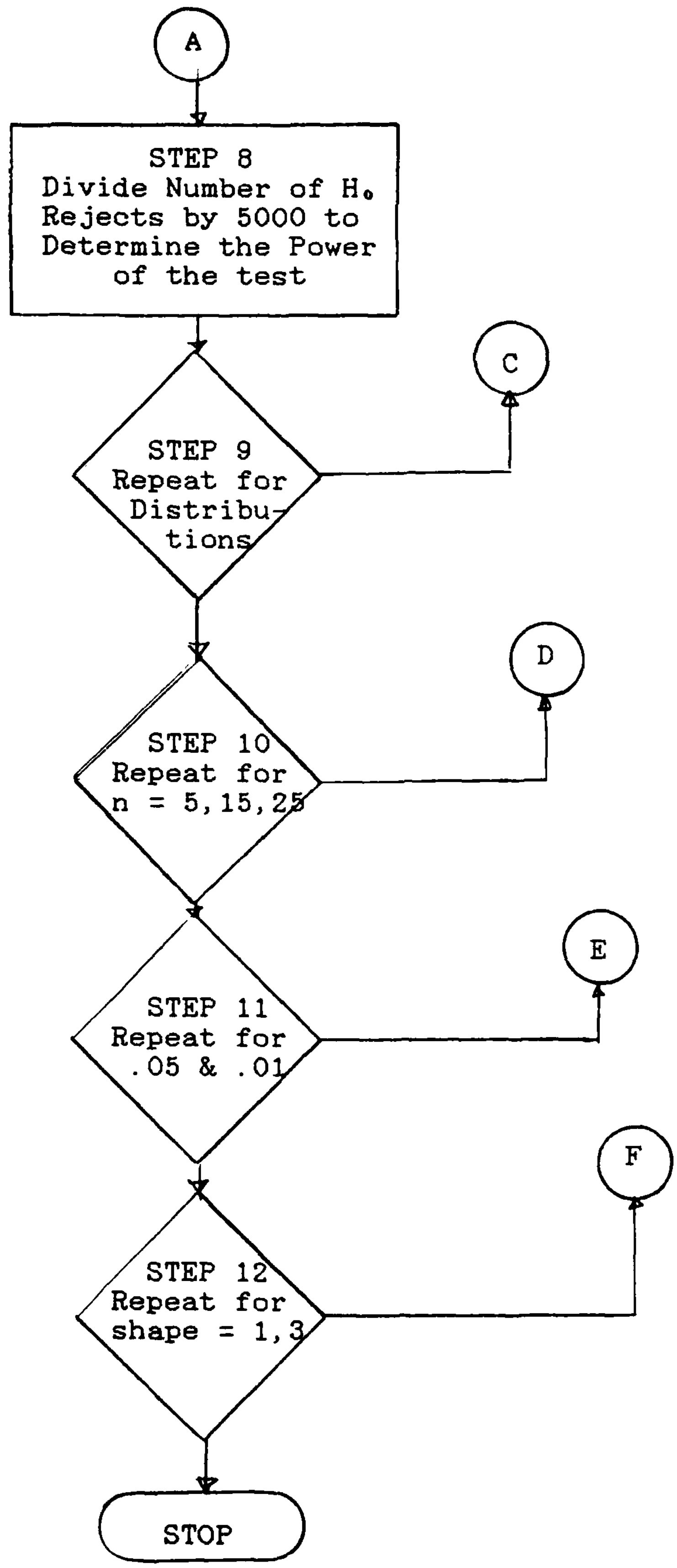


Fig 2 (Continued). Flow Chart for Program POWER

```
****************************************************************
****************************************************************
*                        PROGRAM POWER                         *
****************************************************************
****************************************************************
*
*  PURPOSE: TO TEST THE NULL HYPOTHESIS THAT A SET OF SAMPLE
*           DATA FOLLOWS THE LOGNORMAL DISTRIBUTION WITH
*           SHAPE C AGAINST THE ALTERNATE HYPOTHESIS THAT
*           THE DATA FOLLOWS SOME OTHER DISTRIBUTION.
*
*  VARIABLES:
*     DSEED = RANDOM NUMBER SEED
*         X = RANDOM LOGNORMAL DEVIATES
*         N = SAMPLE SIZE
*         C = SHAPE PARAMETER
*        NC = SAMPLE SIZE * SHAPE
*      AMLE = MLE OF LOCATION PARAMETER
*      BMLE = MLE OF SCALE PARAMETER
*         P = ARRAY OF N POINTS FROM HYPOTHSIZED CDF
*        KS = ARRAY OF VALUES OF MOD. K-S TEST STATISTICS
*        AD = ARRAY OF VALUES OF MOD. A-D TEST STATISTICS
*       CVM = ARRAY OF VALUES OF MOD. C-VM TEST STATISTICS
*        X2 = ARRAY OF VALUES OF CHI-SQUARE TEST STATISTICS
*        IT = ITERATION COUNTER (5000 USED)
*      NSIZ = SAMPLE SIZE COUNTER (1=5,2=15,3=25)
*      NSHP = NULL-HYPOTHESIS LOGNORMAL SHAPE COUNTER (1=1,2=3)
*      NREP = NUMBER OF REPETITIONS TO BE USED
*      NALT = ALTERNATIVE DISTRIBUTION COUNTER
*      NALF = SIGNIFICANT LEVEL COUNTER (1=.05,1=.01)
*      NRKS = NUMBER OF HYPOTHESIS REJECTS UNDER THE K-S TEST
*      NRAD = NUMBER OF HYPOTHESIS REJECTS UNDER THE A-D TEST
*      NRCV = NUMBER OF HYPOTHESIS REJECTS UNDER THE C-VM TEST
*      NRX2 = NUMBER OF HYPOTHESIS REJECTS UNDER THE CHI-2 TEST
*
*  SUBROUTINES:
*    LOGDEV - GENERATES N RANDOM LOGNORMAL DEVIATES
*       MLE - ESTIMATES THE LOCATION AND SCALE PARAMETERS
*      CALC - PREFORMS THE CALCULATIONS FOR THE BI-SECTION
*                        SEARCH USED IN MLE
*    HYPCDF - COMPUTES THE HYPOTHESIZED CDF
*    TESTAT - CALCULATES THE K-S, A-D, AND C-VM TEST STATISTICS
*    COMPAR - COMPARES TEST STATISTICS WITH CRITICAL VALUES AND
*                        COUNTS REJECTS
*
*  IMSL SUBROUTINES:
*     GGNLG - GENERATES LOGNORMAL DEVIATES
*     GGWIB - GENERATES WEIBULL DEVIATES
*     GGAMR - GENERATES GAMMA DEVIATES
*     GGBTR - GENERATES BETA DEVIATES
*     GGEXN - GENERATES EXPONENTIAL DEVIATES
*     GGNML - GENERATES NORMAL DEVIATES
*     VSRTA - ORDERS DATA IN ASCENDING ORDER
*     MDNOR - CALCULATES NORMAL PDF OF VALUE
```

```
*
*          ** NOTE **
*
*  IT IS IMPORTANT TO LINK TO IMSL LIBRARY BEFORE RUNNING THIS
*  PROGRAM
*
***************************************************************
C
      PROGRAM POWER
C
      COMMON  DSEED,X,N,C,NC,AMLE,BMLE,P,
     1          KS,AD,CVM,IT,NSIZ,NSHP,NREP,
     1          NALT,NALF,NRKS,NRAD,NRCV,NRX2,X2
      INTEGER N,NSIZ,NSHP,IT,NREP,NRKS(2,2,3,8),NRAD(2,2,3,8),
     1          NRCV(2,2,3,8),NRX2(2,2,3,8)
      REAL    X(26),AMLE,BMLE,KS(2,2,3,8),AD(2,2,3,8),
     1          CVM(2,2,3,8),C,NC,
     1          P(25),ALPHA,KSPWR(2,2,3,8),ADPWR(2,2,3,8),
     1          CVPWR(2,2,3,8),X2CRIT(2,2,3),X2(2,2,3,8),
     1          X2PWR(2,2,3,8)
C
      CHARACTER TEST(4)*3,ALTCDF(8)*12
      DOUBLE PRECISION DSEED
C
      CALL FILL
C
      TEST(1) = 'K-S'
      TEST(2) = 'A-D'
      TEST(3) = 'CVM'
      TEST(4) = 'CHI'
C
      ALTCDF(1) = 'LOGNORMAL C=1.0'
      ALTCDF(2) = 'LOGNORMAL C=3.0'
      ALTCDF(3) = 'LOGNORMAL C=2.0'
      ALTCDF(4) = 'WEIBULL'
      ALTCDF(5) = 'GAMMA'
      ALTCDF(6) = 'BETA'
      ALTCDF(7) = 'EXPONENTIAL'
      ALTCDF(8) = 'NORMAL'
C
***   OPEN OUTPUT FILE TO STORE COMPUTED POWER VALUES:   ***
C
      OPEN (UNIT=7,FILE='POWER',STATUS='NEW')
C
***   NUMBER OF REPETITIONS TO BE USED ON EACH RUN:   ***

C      PRINT*,'THE MONTE CARLO POWER ANALYSIS WILL REQUIRE'
C      PRINT*,'    5000 REPETITIONS.'
C      PRINT*,'ENTER THE NUMBER TO BE USED FOR THIS RUN:'
C      READ*,NREP
C
      NREP = 5000
C
C      PRINT*,'ENTER RANDOM NUMBER SEED OR "1." FOR DEFAULT:'
```

```
C       READ*,DSEED
C             IF (DSEED .EQ. 1.) DSEED = 123457.0D0
C       PRINT*,' '
C       PRINT*,'STANDBY . . . COMPUTATIONS IN PROGRESS'
C
      DSEED = 123457.0D0
C
      DO 90 NSHP = 1,2
         IF (NSHP .EQ. 1) THEN
             C = 1.0
             WRITE(7,51)
             WRITE(7,56)
             WRITE(7,58)
             WRITE(7,62)
         ELSE IF (NSHP .EQ. 2) THEN
             C = 3.0
             WRITE(7,52)
             WRITE(7,56)
             WRITE(7,59)
             WRITE(7,62)
         END IF
C
         DO 80 NALF = 1,2
C
            IF (NALF .EQ. 1) THEN
               ALPHA = .05
               WRITE(7,64)
            ELSE IF (NALF .EQ. 2) THEN
               ALPHA = .01
               WRITE(7,66)
            END IF
C
            WRITE(7,54)
            WRITE(7,74)
            WRITE(7,68)
            WRITE(7,72)
            WRITE(7,76)
            WRITE(7,72)
C
            NSIZ = 0
C
            DO 70 N = 5,25,10
C
               NSIZ = NSIZ + 1
               NC = N*C

               DO 60 NALT = 1,8

                  NRKS(NSHP,NALF,NSIZ,NALT) = 0
                  NRAD(NSHP,NALF,NSIZ,NALT) = 0
                  NRCV(NSHP,NALF,NSIZ,NALT) = 0
                  NRX2(NSHP,NALF,NSIZ,NALT) = 0
```

```
                 DO 40 IT = 1,NREP
C
                    IF (NALT .EQ. 1) CALL LOGDEV
                    IF (NALT .EQ. 2) CALL LOGDEV
                    IF (NALT .EQ. 3) CALL LOGDEV
                    IF (NALT .EQ. 4) CALL GGWIB(DSEED,3.5,N,X)
                    IF (NALT .EQ. 5)
     1                 CALL GGAMR(DSEED,2.,N,1,X)
                    IF (NALT .EQ. 6)
     1                 CALL GGBTR(DSEED,2.,3.,N,X)
                    IF (NALT .EQ. 7) CALL GGEXN(DSEED,2.,N,X)
                    IF (NALT .EQ. 8) CALL GGNML(DSEED,N,X)
C
                    CALL VSRTA(X,N)
C
                    CALL MLE
C
                    CALL HYPCDF
C
                    CALL TESTAT
C
                    CALL COMPAR
C
 40              CONTINUE
C
                 KSPWR(NSHP,NALF,NSIZ,NALT)
     1              = NRKS(NSHP,NALF,NSIZ,NALT)/REAL(NREP)
                 ADPWR(NSHP,NALF,NSIZ,NALT)
     1              = NRAD(NSHP,NALF,NSIZ,NALT)/REAL(NREP)
                 CVPWR(NSHP,NALF,NSIZ,NALT)
     1              = NRCV(NSHP,NALF,NSIZ,NALT)/REAL(NREP)
                 X2PWR(NSHP,NALF,NSIZ,NALT)
     1              = NRX2(NSHP,NALF,NSIZ,NALT)/REAL(NREP)
C
 60           CONTINUE
C
***           WRITE POWER RESULTS TO FILE   ***
C
              WRITE(7,110),N,TEST(1),KSPWR(NSHP,NALF,NSIZ,1),
     1        KSPWR(NSHP,NALF,NSIZ,2),KSPWR(NSHP,NALF,NSIZ,3),
     1        KSPWR(NSHP,NALF,NSIZ,4),KSPWR(NSHP,NALF,NSIZ,5),
     1        KSPWR(NSHP,NALF,NSIZ,6),KSPWR(NSHP,NALF,NSIZ,7),
     1        KSPWR(NSHP,NALF,NSIZ,8)
C
              WRITE(7,110),N,TEST(2),ADPWR(NSHP,NALF,NSIZ,1),
     1        ADPWR(NSHP,NALF,NSIZ,2),ADPWR(NSHP,NALF,NSIZ,3),
     1        ADPWR(NSHP,NALF,NSIZ,4),ADPWR(NSHP,NALF,NSIZ,5),
     1        ADPWR(NSHP,NALF,NSIZ,6),ADPWR(NSHP,NALF,NSIZ,7),
     1        ADPWR(NSHP,NALF,NSIZ,8)
C
              WRITE(7,110),N,TEST(3),CVPWR(NSHP,NALF,NSIZ,1),
     1        CVPWR(NSHP,NALF,NSIZ,2),CVPWR(NSHP,NALF,NSIZ,3),
     1        CVPWR(NSHP,NALF,NSIZ,4),CVPWR(NSHP,NALF,NSIZ,5),
```

```
     1          CVPWR(NSHP,NALF,NSIZ,6),CVPWR(NSHP,NALF,NSIZ,7),
     1          CVPWR(NSHP,NALF,NSIZ,8)
C
                WRITE(7,110),N,TEST(4),X2PWR(NSHP,NALF,NSIZ,1),
     1          X2PWR(NSHP,NALF,NSIZ,2),X2PWR(NSHP,NALF,NSIZ,3),
     1          X2PWR(NSHP,NALF,NSIZ,4),X2PWR(NSHP,NALF,NSIZ,5),
     1          X2PWR(NSHP,NALF,NSIZ,6),X2PWR(NSHP,NALF,NSIZ,7),
     1          X2PWR(NSHP,NALF,NSIZ,8)
C
                WRITE(7,72)
C
 70          CONTINUE
C
 80       CONTINUE
C
          WRITE(7,74)
C
 90    CONTINUE
C
 51    FORMAT('1',36X,'TABLE IV')
 52    FORMAT('1',35X,'TABLE V')
 54    FORMAT(' ')
 56    FORMAT('0',22X,'POWER TEST FOR THE LOGNORMAL ',
     1       'DISTRIBUTION')
 58    FORMAT(22X,'HO: LOGNORMAL DISTRIBUTION AT SHAPE C = 1.0')
 59    FORMAT(22X,'HO: LOGNORMAL DISTRIBUTION AT SHAPE C = 3.0')
 62    FORMAT(22X,'HA: THE DATA FOLLOW ANOTHER DISTRIBUTION')
 64    FORMAT('0',28X,'LEVEL OF SIGNIFICANCE = .05')
 66    FORMAT('0',28X,'LEVEL OF SIGNIFICANCE = .01')
 68    FORMAT(35X,'ALTERNATE DISTRIBUTIONS')
 72    FORMAT(80('-'))
 74    FORMAT(80('='))
 76    FORMAT(2X,' N',3X,'TEST',4X,'PAR.1',3X,'PAR.2',3X,
     1         'PAR.3',3X,'WEIBL',3X,'GAMMA',3X,'BETA',4X,
     1         'EXPON',3X,'NORML')
C
 110    FORMAT(' ',I3,A7,F9.3,7F8.3)
C
        CLOSE(7)
C
        END
C
***    END MAIN PROGRAM   ***
C
```

```
****************************************************************
*
*  PURPOSE:  TO FILL ALL ARRAYS USED IN THIS PROGRAM WITH THE
*            VALUE OF 0
*
*  VARIABLES:
*          X = ARRAY OF RANDOM LOGNORMAL DEVIATES
*          P = ARRAY OF N POINTS FROM HYPOTHSIZED CDF
*         KS = ARRAY OF VALUES OF MODIFIED K-S TEST STATISTICS
*         AD = ARRAY OF VALUES OF MODIFIED A-D TEST STATISTICS
*        CVM = ARRAY OF VALUES OF MODIFIED C-VM TEST STATISTICS
*         X2 = ARRAY OF VALUES OF CHI-SQUARE TEST STATISTICS
*       NRKS = ARRAY OF REJECTION OF THE K-S TEST
*       NRAD = ARRAY OF REJECTION OF THE A-D TEST
*       NRCV = ARRAY OF REJECTION OF THE C-VM TEST
*       NRX2 = ARRAY OF REJECTION OF THE CHI-SQUARE TEST
*      KSPWR = ARRAY OF POWERS OF THE MODIFIED K-S TEST
*      ADPWR = ARRAY OF POWERS OF THE MODIFIED A-D TEST
*      CVPWR = ARRAY OF POWERS OF THE MODIFIED C-VM TEST
*      X2PWR = ARRAY OF POWERS OF THE CHI-SQUARE TEST
*     X2CRIT = ARRAY OF CRITICAL VALUES FOR CHI-SQUARE TEST
*
****************************************************************
C
       SUBROUTINE FILL
C
      COMMON  DSEED,X,N,C,NC,AMLE,BMLE,P,
     1          KS,AD,CVM,IT,NSIZ,NSHP,NREP,
     1          NALT,NALF,NRKS,NRAD,NRCV,NRX2,X2
      INTEGER N,NSIZ,NSHP,IT,NREP,NRKS(2,2,3,8),NRAD(2,2,3,8),
     1          NRCV(2,2,3,8),NRX2(2,2,3,8)
      REAL    X(26),AMLE,BMLE,KS(2,2,3,8),AD(2,2,3,8),
     1          CVM(2,2,3,8),C,NC,
     1          P(25),ALPHA,KSPWR(2,2,3,8),ADPWR(2,2,3,8),
     1          CVPWR(2,2,3,8),X2CRIT(2,2,3),X2(2,2,3,8),
     1          X2PWR(2,2,3,8)
C
      DOUBLE PRECISION DSEED
C
      DO 10 I=1,25
C
          X(I) = 0.0
          P(I) = 0.0
C
 10   CONTINUE
C
      DO 20 I=1,2
C
        DO 30 J=1,2
C
          DO 40 K=1,3
C
```

```
                DO 50 L=1,8
                  NRKS(I,J,K,L)=0
                  NRAD(I,J,K,L)=0
                  NRCV(I,J,K,L)=0
                  NRX2(I,J,K,L)=0
                  KS(I,J,K,L)=0.0
                  AD(I,J,K,L)=0.0
                  CVM(I,J,K,L)=0.0
                  X2(I,J,K,L)=0.0
                  KSPWR(I,J,K,L)=0.0
                  ADPWR(I,J,K,L)=0.0
                  CVPWR(I,J,K,L)=0.0
                  X2PWR(I,J,K,L)=0.0
C
 50             CONTINUE
C
                X2CRIT(I,J,K)=0.0
C
 40           CONTINUE
C
 30         CONTINUE
C
 20       CONTINUE
C
          RETURN
          END
C
*** END SUBROUTINE FILL
C
```

```
*****************************************************************
*
*  PURPOSE:  TO GENERATE N RANDOM DEVIATES FROM A LOGNORMAL
*            DISTRIBUTION WHOSE PARENT NORMAL HAS A MEAN
*            OF 0 AND A STANDARD DEVIATION OF 1, THE SUBROUTINE
*            THEN ADDS THE LOCATION OF 10 TO EACH OF THE
*            DEVIATES.  VRSTA THEN ORDERS THE SAMPLE DATA.
*
*  VARIABLES:
*          DSEED = RANDOM NUMBER SEED
*              X = RANDOM LOGNORMAL DEVIATES
*              N = SAMPLE SIZE
*            PMU = MEAN OF PARENT NORMAL OF LOGNORMAL
*           PVAR = VARIANCE OF PARENT NORMAL OF LOGNORMAL
*
*  IMSL SUBROUTINES:
*          GGNLG - GENERATES LOGNORMAL RANDOM DEVIATES
*          VSRTA - ORDERS DATA IN ACENDING ORDER
*
*****************************************************************
C
      SUBROUTINE LOGDEV
C
      COMMON  DSEED,X,N,C,NC,AMLE,BMLE,P,
     1          KS,AD,CVM,IT,NSIZ,NSHP,NREP,
     1          NALT,NALF,NRKS,NRAD,NRCV,NRX2,X2
      INTEGER N,NSIZ,IT,NREP,NRKS(2,2,3,8),NRAD(2,2,3,8),
     1          NRCV(2,2,3,8),NRX2(2,2,3,8)
      REAL    X(26),AMLE,BMLE,KS(2,2,3,8),AD(2,2,3,8),
     1          CVM(2,2,3,8),C,NC,
     1          P(25),ALPHA,KSPWR(2,2,3,8),ADPWR(2,2,3,8),
     1          CVPWR(2,2,3,8),X2CRIT(2,2,3),X2(2,2,3,8),
     1          X2PWR(2,2,3,8),AC
      DOUBLE PRECISION DSEED
C
      PMU  = 0.0
      PVAR = 1.0
C
      CALL GGNLG(DSEED,N,PMU,PVAR,X)
C
       DO 10 I=1,N
          X(I) = X(I) + 10.0
   10    CONTINUE
C
      CALL VSRTA(X,N)
C
      RETURN
C
      END
C
***   END SUBROUTINE LOGDEV   ***
C
```

```
*****************************************************************
*
*  PURPOSE:  TO ESTIMATE THE MAXIMUM LIKELIHOOD ESTIMATORS OF
*            THE LOCATION AND THE SCALE PARAMETERS - GIVEN A
*            KNOWN SHAPE PARAMETER, C.
*
*  VARIABLES:
*        X = RANDOM LOGNORMAL DEVIATES
*        N = SAMPLE SIZE
*        C = SHAPE PARAMETER
*     AMLE = MAXIMUM LIKELIHOOD ESTIMATOR OF LOCATION PARAMETER
*     BMLE = MAXIMUM LIKELIHOOD ESTIMATOR OF SCALE PARAMETER
*      DIF = VARIABLE USED IN CALCULATIONS
*     TDIF = VARIABLE USED IN CALCULATIONS
*     TEMP = VARIABLE USED IN CALCULATIONS
*       UP = UPPER BOUND OF LOCATION PARAMETER
*    UPPER = VALUE RETURNED BY CALC FOR UP
*      LOW = LOWER OF STEPS IN BISECTION SEARCH
*    LOWER = VALUE RETURNED BY CALC FOR LOW
*      MID = VALUE OF THE MID-POINT BETWEEN LOW AND UP
*   MIDDLE = VALUE RETURNED BY CALC FOR MID
*     STEP = SIZE OF BACKWARD STEP = 10% OF X(1)
*    THETA = VARIABLE USED IN CALCULATIONS
*   STHETA = SUM OF ALL THETA
*   HTHETA = LOGNORMAL OF THE ESTIMATE FOR THE SCALE PARAMETER
*
*  SUBROUTINES:
*       CALC - PERFORMS CALCULATIONS FOR THE BISECTION SEARCH
*
*****************************************************************
C
      SUBROUTINE MLE
C
      COMMON  DSEED,X,N,C,NC,AMLE,BMLE,P,
     1          KS,AD,CVM,IT,NSIZ,NSHP,NREP,
     1          NALT,NALF,NRKS,NRAD,NRCV,NRX2,X2
      INTEGER N,NSIZ,NSHP,IT,NREP,NRKS(2,2,3,8),NRAD(2,2,3,8),
     1          NRCV(2,2,3,8),NRX2(2,2,3,8)
      REAL    X(26),AMLE,BMLE,KS(2,2,3,8),AD(2,2,3,8),
     1          CVM(2,2,3,8),C,NC,
     1          P(25),ALPHA,KSPWR(2,2,3,8),ADPWR(2,2,3,8),
     1          CVPWR(2,2,3,8),X2CRIT(2,2,3),X2(2,2,3,8),
     1          X2PWR(2,2,3,8)
      DOUBLE PRECISION DSEED
C
      REAL LOW,LOWER,MID,MIDDLE,UP,UPPER,STEP,THETA,STHETA,
     1           HTHETA
C
      UP = X(1)
C
      CALL CALC(UP,X,N,C,UPPER)
C
      STEP = (ABS(.1*X(1)))
```

```
C
      LOW = X(1) - STEP
C
 5    CONTINUE
C
      CALL CALC(LOW,X,N,C,LOWER)
C
      IF ((UPPER*LOWER) .GT. 0.0) THEN
         UP = LOW
         LOW = LOW - STEP
         UPPER = LOWER
         GO TO 5
      END IF
C
 10   CONTINUE
C
      MID = (UP + LOW)/2
C
      CALL CALC(MID,X,N,C,MIDDLE)
C
      IF ((UPPER*MIDDLE) .LE. 0.0) THEN
         LOW = MID
         LOWER = MIDDLE
      ELSE
         UP = MID
         UPPER = MIDDLE
      END IF
C
      IF (ABS(UP-LOW) .GT. .01) THEN
         GO TO 10
      END IF
C
      AMLE = MID
C
      STHETA = 0.0
C
      DO 15 I=1,N
         TEMP = LOG(X(I) - AMLE)
         STHETA = STHETA + TEMP
 15   CONTINUE
C
      HTHETA = STHETA/N
      BMLE = HTHETA
C
      RETURN
C
      END
C
***   END SUBROUTINE MLE   ***
C
```

```
*****************************************************************
*
*  PURPOSE:  PERFORM THE CALCULATION NEEDED TO TRANSFORM THE
*            ESTIMATED LOGNORMAL PARAMETERS TO THE PARENT
*            PARAMETER.
*
*  VARIABLES:
*          X = RANDOM LOGNORMAL DEVIATES
*          N = SAMPLE SIZE
*        SHP = SHAPE PARAMETER
*        LOC = CURRENT LOCATION PARAMETER USED IN CALCULATIONS
*       TSUM = VARIABLE USED IN CALCULATIONS
*        DIF = VARIABLE USED IN CALCULATIONS
*       TDIF = VARIABLE USED IN CALCULATIONS
*        SUM = VARIABLE USED IN CALCULATIONS
*      LNDIF = VARIABLE USED IN CALCULATIONS
*
*****************************************************************
C
      SUBROUTINE CALC(LOC,X,N,SHP,TSUM)
C
      INTEGER N
      REAL LOC,X(31),SHP,TSUM,DIF,LNDIF,TDIF,SUM
C
      SUM = 0.0
      TSUM = 0.0
C
      DO 5 I = 1,N
         DIF = X(I) - LOC
         IF (DIF .EQ. 0.0) DIF = .0001
         LNDIF = LOG(DIF)
         SUM = SUM+LNDIF
 5    CONTINUE
C
      SUM = SUM/N
C
      DO 10 I =1,N
         DIF = X(I) - LOC
         IF (DIF .EQ. 0.0) DIF = .0001
         LNDIF = LOG(DIF)
         TDIF = 1/DIF
         TSUM = TSUM+TDIF+(1/SHP)*TDIF*(LNDIF-SUM)
 10   CONTINUE
C
      RETURN
C
      END
C
***   END SUBROUTINE CALC   ***
C
```

```
***************************************************************
*
*  PURPOSE:  TO COMPUTE THE HYPOTHESIZED LOGNORMAL DISTRIBUTION
*            FUNCTION P(I) FOR I = 1,2,...,N - GIVEN A KNOWN
*            SHAPE PARAMETER AND THE ESTIMATED VALUES FOR THE
*            LOCATION AND SCALE PARAMETER.
*
*  VARIABLES:
*        X = RANDOM LOGNORMAL DEVIATES
*        N = SAMPLE SIZE
*        C = SHAPE PARAMETER
*     AMLE = MAXIMUM LIKELIHOOD ESTIMATOR OF LOCATION PARAMETER
*     BMLE = MAXIMUM LIKELIHOOD ESTIMATOR OF SCALE PARAMETER
*        P = ARRAY OF N POINTS OF THE HYPOTHESIZED CDF
*
*  IMSL SUBROUTINES:
*      MDNOR - CALCULATES THE NORMAL PDF OF OBSERVATION
*
***************************************************************
C
      SUBROUTINE HYPCDF
C
      COMMON  DSEED,X,N,C,NC,AMLE,BMLE,P,
     1          KS,AD,CVM,IT,NSIZ,NSHP,NREP,
     1          NALT,NALF,NRKS,NRAD,NRCV,NRX2,X2
      INTEGER N,NSIZ,NSHP,IT,NREP,NRKS(2,2,3,8),NRAD(2,2,3,8),
     1          NRCV(2,2,3,8),NRX2(2,2,3,8)
      REAL    X(26),AMLE,BMLE,KS(2,2,3,8),AD(2,2,3,8),
     1          CVM(2,2,3,8),C,NC,
     1          P(25),ALPHA,KSPWR(2,2,3,8),ADPWR(2,2,3,8),
     1          CVPWR(2,2,3,8),X2CRIT(2,2,3),X2(2,2,3,8),
     1          X2PWR(2,2,3,8)
      DOUBLE PRECISION DSEED
C
      REAL Q,Z
C
      DO 10 I = 1,N
             Q = (LOG(X(I)-AMLE) - BMLE)/C
             CALL MDNOR(Q,Z)
             P(I) = Z
 10      CONTINUE
C
      RETURN
C
      END
C
***   END SUBROUTINE HYPCDF
C
```

```
*****************************************************************
*
*  PURPOSE:  COMPUTE VALUES OF THE TEST STATISTICS OF THE
*            CHI-SQUARE  AND THE MODIFIED K-S, A-D, C-VM
*            GOODNESS-OF-FIT TESTS.
*
*  VARIABLES:
*      X = RANDOM LOGNORMAL DEVIATES
*      N = SAMPLE SIZE
*      C = SHAPE PARAMETER
*   AMLE = MAXIMUM LIKELIHOOD ESTIMATOR OF LOCATION PARAMETER
*   BMLE = MAXIMUM LIKELIHOOD ESTIMATOR OF SCALE PARAMETER
*      P = ARRAY OF N POINTS FROM HYPOTHSIZED CDF
*     KS = ARRAY OF VALUES OF MODIFIED K-S TEST STATISTICS
*     AD = ARRAY OF VALUES OF MODIFIED A-D TEST STATISTICS
*    CVM = ARRAY OF VALUES OF MODIFIED C-VM TEST STATISTICS
*     X2 = ARRAY OF VALUES OF CHI-SQUARE TEST STATISTICS
*     IT = ITERATION COUNTER (5000 USED)
*   NSIZ = SAMPLE SIZE COUNTER (1=5,2=15,3=25)
*   NSHP = NULL-HYPOTHESIS LOGNORMAL SHAPE COUNTER (1=1,2=3)
*   NREP = NUMBER OF REPETITIONS TO BE USED
*   NALT = ALTERNATIVE DISTRIBUTION COUNTER
*   NALF = SIGNIFICANT LEVEL COUNTER (1=.05,1=.01)
*
*     DP = POSITIVE DIFFERENCES BETWEEN EDF AND CDF POINTS
*     DM = NEGATIVE DIFFERENCES BETWEEN EDF AND CDF POINTS
*  DPLUS = MAXIMUM POSITIVE DIFFERENCE (LARGEST DP VALUE)
* DMINUS = MAXIMUM NEGATIVE DIFFERENCE (LARGEST DM VALUE)
*     KS = VALUES OF THE MODIFIED K-S TEST STATISTIC
*
*     AL = VALUE USED TO CALCULATE THE A-D TEST STATISTIC
*     AM = VALUE USED TO CALCULATE THE A-D TEST STATISTIC
*     AN = AL + AM
*    AAA = VALUES TO TO BE SUMMED FOR A-D TEST STATISTIC
*   SAAA = SUM OF AAA VALUE
*     AD = VALUES OF THE MODIFIED A-D TEST STATISTIC
*
*    ACV = SQUARED QUANTITIES IN THE C-VM FORMULA
*   SACV = SUM OF ACV VALUES
*    CVM = VALUES OF THE MODIFIED C-VM TEST STATISTIC
*
*****************************************************************
C
      SUBROUTINE TESTAT
C
      COMMON  DSEED,X,N,C,NC,AMLE,BMLE,P,
     1          KS,AD,CVM,IT,NSIZ,NSHP,NREP,
     1          NALT,NALF,NRKS,NRAD.NRCV,NRX2,X2
      INTEGER N,NSIZ,NSHP,IT,NREP,NRKS(2,2,3,8),NRAD(2,2,3,8),
     1          NRCV(2,2,3,8),OBS(5),NRX2(2,2,3,8)
      REAL    X(26),AMLE,BMLE,KS(2,2,3,8),AD(2,2,3,8),
     1          CVM(2,2,3,8),C,NC,
     1          P(25),ALPHA,KSPWR(2,2,3,8),ADPWR(2,2,3,8),
```

```
     1           CVPWR(2,2,3,8),X2CRIT(2,2,3),X2(2,2,3,8),
     1           EX,DP(25),DM(25),DPLUS,DMINUS,AL(25),
     1           X2PWR(2,2,3,8),AM(25),AN(25),AAA(25),SAAA,
     1           ACV(25),SACV,RTEND(4)
      DOUBLE PRECISION DSEED
C
      DPLUS = 0
      DMINUS = 0
      DO 5 IK = 1,25
         DP(IK) = 0
         DM(IK) = 0
 5    CONTINUE
C
      DO 10 I = 1,N
         DP(I) = ABS((I/REAL(N)) - P(I))
         DM(I) = ABS(P(I) - (I-1)/REAL(N))
 10    CONTINUE
C
      DPLUS = MAX( DP(1),DP(2),DP(3),DP(4),DP(5),DP(6),DP(7),
     1       DP(8),DP(9),DP(10),DP(11),DP(12),DP(13),DP(14),
     1       DP(15),DP(16),DP(17),DP(18),DP(19),DP(20),
     1       DP(21),DP(22),DP(23),DP(24),DP(25) )
C
      DMINUS = MAX( DM(1),DM(2),DM(3),DM(4),DM(5),DM(6),DM(7),
     1       DM(8),DM(9),DM(10),DM(11),DM(12),DM(13),DM(14),
     1       DM(15),DM(16),DM(17),DM(18),DM(19),DM(20),
     1       DM(21),DM(22),DM(23),DM(24),DM(25) )
C
       KS(NSHP,NALF,NSIZ,NALT) = MAX(DPLUS,DMINUS)
C
       SAAA = 0
C
       DO 20 J= 1,N
           IF (P(J) .LE. .001) P(J) = .001
           AL(J) = LOG (P(J))
           IF (P(N+1-J) .LE. .001) P(N+1-J) = .001
           AM(J) = LOG (1.0 - P(N+1-J))
           AN(J) = AL(J) + AM(J)
           AAA(J) = (2.0*J - 1.0) * AN(J)
           SAAA = SAAA + AAA(J)
 20   CONTINUE
C
      AD(NSHP,NALF,NSIZ,NALT) = -N - (1.0/REAL(N)) * SAAA
C
      SACV = 0
C
      DO 30 K = 1,N
          ACV(K) = (P(K) - (2.0*K - 1.0)/(2.0*REAL(N)))**2
          SACV = SACV + ACV(K)
 30   CONTINUE
C
      CVM(NSHP,NALF,NSIZ,NALT) = SACV + (1.0/( 12.0*REAL(N)))
C
```

```
      DO 40 IN = 1,5
          OBS(IN) = 0
 40   CONTINUE
C
      DO 50 KI = 1,4
          RTEND(KI) = AMLE-BMLE + BMLE*(1.0-.2*KI)**(-1.0/C)
 50   CONTINUE
C
      DO 60 M = 1,N
C
         IF (X(M) .LE. RTEND(1) ) THEN
            OBS(1) = OBS(1) + 1
         ELSE IF (X(M) .LE. RTEND (2)) THEN
            OBS(2) = OBS(2) +1
         ELSE IF (X(M).LE.RTEND(3)) THEN
            OBS(3) = OBS(3) + 1
         ELSE IF (X(M) .LE. RTEND(4)) THEN
            OBS(4) = OBS(4) + 1
         ELSE
            OBS(5) = OBS(5) + 1
         END IF
C
 60   CONTINUE
C
      EX = N/5.0
C
      X2(NSHP,NALF,NSIZ,NALT) = ( (OBS(1)-EX) **2)/ EX
     1       + ((OBS(2)-EX)**2)/EX + ((OBS(3)-EX)**2)/EX
     1       + ((OBS(4)-EX)**2)/EX + ((OBS(5)-EX)**2)/EX
C
      RETURN
C
      END
C
***   END SUBROUTINE TESTAT   ***
C
```

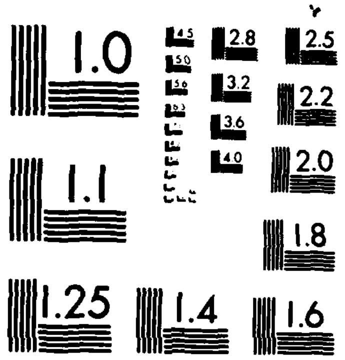
```
****************************************************************
*
*  PURPOSE:  COMPARE THE TEST STATISTIC CALCULATED FROM THE
*            CHI-SQUARE OR ONE OF THE MODIFIED K-S, A-D, C-VM
*            TESTS, WITH CORRECT VALUES (THE MODIFIED TEST ARE
*            FROM PROGRAM CRITICAL) AND COUNT THE NUMBER OF
*            TIMES THE NULL HYPOTHESIS IS  REJECTED.
*
*  VARIABLES:
*           X = RANDOM LOGNORMAL DEVIATES
*           N = SAMPLE SIZE
*          KS = ARRAY OF VALUES OF MODIFIED K-S TEST STATISTICS
*          AD = ARRAY OF VALUES OF MODIFIED A-D TEST STATISTICS
*         CVM = ARRAY OF VALUES OF MODIFIED C-VM TEST STATISTICS
*          X2 = ARRAY OF VALUES OF CHI-SQUARE TEST STATISTICS
*        NSIZ = SAMPLE SIZE COUNTER (1=5,2=15,3=25)
*        NSHP = NULL-HYPOTHESIS LOGNORMAL SHAPE COUNTER (1=1,2=3)
*        NREP = NUMBER OF REPETITIONS TO BE USED
*        NALT = ALTERNATIVE DISTRIBUTION COUNTER
*        NALF = SIGNIFICANT LEVEL COUNTER (1=.05,1=.01)
*        NRKS = NUMBER OF HYPOTHESIS REJECTS UNDER THE K-S TEST
*        NRAD = NUMBER OF HYPOTHESIS REJECTS UNDER THE A-D TEST
*        NRCV = NUMBER OF HYPOTHESIS REJECTS UNDER THE C-VM TEST
*        NRX2 = NUMBER OF HYPOTHESIS REJECTS UNDER THE CHI 2 TEST
*      KSCRIT = ARRAY OF MODIFIED CRITICAL VALUES
*      ADCRIT = ARRAY OF MODIFIED CRITICAL VALUES
*      CVCRIT = ARRAY OF MODIFIED CRITICAL VALUES
*      X2CRIT = ARRAY OF CHI-SQUARE CRITICAL VALUES
*
****************************************************************
C
      SUBROUTINE COMPAR
C
      COMMON  DSEED,X,N,C,NC,AMLE,BMLE,P,
     1          KS,AD,CVM,IT,NSIZ,NSHP,NREP
     1          NALT,NALF,NRKS,NRAD,NRCV,NRX
      INTEGER N,NSIZ,NSHP,IT,NREP,NRKS
     1          NRCV(2,2,3,8),NRX2(2,
      REAL    X(26),AMLE,BMLE,KS(2,
     1          CVM(2,2,3,8),C,NC
     1          P(25),ALPHA,KSPWR
     1          CVPWR(2,2,3,8),KSCRIT
     1          CVCRIT(2,2,3 X2CRIT
     1          X2PWR(2,2,3,8
      DOUBLE PRECISION DSEED
C
***   INPUT K-S CRITICAL VALUES
C
      KSCRIT(1,1,1
      KSCRIT(1,1,2
      KSCRIT(1,1,3
      KSCRIT(1,2,1
      KSCRIT(1,
```

AD-A179 049 A MODIFIED GOODNESS-OF-FIT TEST FOR THE LOGNORMAL DISTRIBUTION WITH UNKNO..(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGI.. L I WHITSEL

2/2

UNCLASSIFIED DEC 86 AFIT/GOR/MA/86D-6 F/G 12/1 NL

END
DATE
FILMED
5-87



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

```
      KSCRIT(1,2,3) = .5522
      KSCRIT(2,1,1) = .5544
      KSCRIT(2,1,2) = .4622
      KSCRIT(2,1,3) = .4380
      KSCRIT(2,2,1) = .6148
      KSCRIT(2,2,2) = .4891
      KSCRIT(2,2,3) = .4635
C
***   INPUT A-D CRITICAL VALUES FROM TABLE VII:   ***
      ADCRIT(1,1,1) = 7.2321
      ADCRIT(1,1,2) = 10.7748
      ADCRIT(1,1,3) = 15.2449
      ADCRIT(1,2,1) = 10.9389
      ADCRIT(1,2,2) = 12.9161
      ADCRIT(1,2,3) = 17.3529
      ADCRIT(2,1,1) = 1.8169
      ADCRIT(2,1,2) = 3.5184
      ADCRIT(2,1,3) = 5.6121
      ADCRIT(2,2,1) = 2.3933
      ADCRIT(2,2,2) = 3.8479
      ADCRIT(2,2,3) = 6.0957
C
***   INPUT C-VM CRITICAL VALUES FROM TABLE VIII:   ***
C
      CVCRIT(1,1,1) = .8858
      CVCRIT(1,1,2) = 1.6899
      CVCRIT(1,1,3) = 2.5205
      CVCRIT(1,2,1) = 1.2142
      CVCRIT(1,2,2) = 1.9970
      CVCRIT(1,2,3) = 2.7992
      CVCRIT(2,1,1) = .3793
      CVCRIT(2,1,2) = .7572
      CVCRIT(2,1,3) = 1.1950
      CVCRIT(2,2,1) = .4963
      CVCRIT(2,2,2) = .8336
      CVCRIT(2,2,3) = 1.3053
C
***   INPUT CHI-SQUARE CRITICAL VALUES :   ***
C
      X2CRIT(1,1,1) = 6.000003
      X2CRIT(1,1,2) = 7.333337
      X2CRIT(1,1,3) = 7.600005
      X2CRIT(1,2,1) = 12.00000
      X2CRIT(1,2,2) = 10.66667
      X2CRIT(1,2,3) = 10.80000
      X2CRIT(2,1,1) = 6.000003
      X2CRIT(2,1,2) = 7.333337
      X2CRIT(2,1,3) = 7.600005
      X2CRIT(2,2,1) = 6.000003
      X2CRIT(2,2,2) = 10.46378
      X2CRIT(2,2,3) = 10.80000
C
```

```
      IF (KS(NSHP,NALF,NSIZ,NALT) .GT. KSCRIT(NSHP,NALF,NSIZ))
     1 NRKS(NSHP,NALF,NSIZ,NALT)=NRKS(NSHP,NALF,NSIZ,NALT) + 1
C
      IF (AD(NSHP,NALF,NSIZ,NALT) .GT. ADCRIT(NSHP,NALF,NSIZ))
     1 NRAD(NSHP,NALF,NSIZ,NALT)=NRAD(NSHP,NALF,NSIZ,NALT) + 1
C
      IF (CVM(NSHP,NALF,NSIZ,NALT).GT.CVCRIT(NSHP,NALF,NSIZ))
     1 NRCV(NSHP,NALF,NSIZ,NALT)=NRCV(NSHP,NALF,NSIZ,NALT) + 1
C
      IF ( X2(NSHP,NALF,NSIZ,NALT).GT.X2CRIT(NSHP,NALF,NSIZ))
     1 NRX2(NSHP,NALF,NSIZ,NALT)=NRX2(NSHP,NALF,NSIZ,NALT) + 1
C
      RETURN
C
      END
C
***   END SUBROUTINE COMPAR   ***
```

BIBLIOGRAPHY

1. Aitchison, J. and J. A. C. Brown. The Lognormal Distribution. Cambridge: University Press, 1957.

2. Anderson, T. W. and D. A. Darling. "A Test of Goodness of Fit," Journal of the American Statistical Association, 49: 765-769 (Dec 1954).

3. Banks, Jerry and John S. Carson. Discrete-Event System Simulation. Englewood Cliffs: Prentice-Hall, 1984.

4. Bush, J. G. and others. "Modified Cramer-von Mises and Anderson-Darling Tests for Weibull Distributions with Unknown Location and Scale Parameters," Communications in Statistics, Part A- Theory and Methods, 12: 240-245 (1983).

5. Buslendo, N. P., and others. The Monte Carlo Method. New York: Pergamon Press, 1966.

6. Cohen, A. C. "Estimating the Parameters of Logarithmic-Normal Distributions by Maximum Likelihood," Journal of the American Statistical Association, 46: 206-212 (1951).

7. Conover, W. J. Practical Nonparametric Statistics (Second Edition). New York: John Wiley and Sons, 1980.

8. David, F. N. and N. L. Johnson. "The Probability Integral Transformation when Parameters are Estimated from the Sample," Biometrika, 35: 182-190 (1948).

9. Deutsch, Ralph. Estimation Theory. Englewood Cliffs NJ: Prentice-Hall, Inc., 1965.

10. Easterling, Robert G. "Goodness of Fit and Parameter Estimatiton," Technometrics, 18: 1-9 (Feb 1976).

11. Green, J. and Y. Hegazy. "Powerful Modified EDF Goodness-of-Fit Tests," Journal of the American Statistical Association, 71: 204-209 (1976).

12. Hammersley, J. M. and D. C. Handcomb. Monte Carlo Methods. London: Methuen and Co., 1967.

13. Harter, H. L. and A. H. Moore. "Local-Maximum Likeilhood Estimation of the Parameters of Three-Parameter Lognormal Populations from Complete and Censored Samples," Journal of the American Statistical Association, 61: 842-851 (1966).

14. Hill, B. M. "The Three-Parameter Lognorma Distributiton and Bayesian Analysis of the Point Source Epidemic," Journal of the American Statistical Association, 58: 72-84 (1963).

15. IMSL Inc. IMSL Library Reference Manual. Houston: IMSL Inc., 1980.

16. Johnson, N. L. "Systems of Frequency Curves Generated by Methods of Translation," Biometrika, 36: 149-176 (1949).

17. Kapur, D.C. and L.R. Lamberson. Reliability in Engineering Design. New York: John Wiley and Sons, 1977.

18. Keefer, Jim H. Robust Minimum Distance Estimation of the Three Parameter Lognormal Distribution. MS Thesis, GOR/MA/83D-3. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, Dec 1983.

19. Lariccia, Vincent N., Ross P. Kindermann. "An Asymptotically Efficient Closed Form Estimator for the Three-Parameter Lognormal Distribution", Communications in Statistics, Part A- Theory and Methods, 12: 243-261 (1983).

20. Lilliefors, Hubert W. "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown", Journal of the American Statistical Association, 62: 399-402 (1967).

21. Lilliefors, Hubert W. "On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown." Journal of the American Statistical Association, 64:387-399 (1969).

22. Mann, N. R., E. M. Scheuer, and K. W. Fertig. "A New Goodness-of-Fit Test for the Two-Parameter Weibull or Extreme-Value Distribution with Unknown Parameters", Communications in Statistics, 2: 383-400 (1973).

23. Massey, Frank J. "The Kolmogorov-Smirnov Test for Goodness of Fit", Journal of the American Statistical Association, 46: 68-78 (1951).

24. Monlezun, Charles J., Lawrence A. Klimko. "Estimating Location and Scale Parameters in the Three Parameter Lognormal with Known Shape Parameter," Presented at the Thirty Sixth Annual Meeting of the Institute of Mathematical Statistics in New York City, Friday, December 28, 1973.

25. Mood, M. M., F. A. Graybill, and D. C. Boes. Introduction to the Theory of Statistics (Third Edition). New York: McGraw-Hill Book Company, 1974.

26. Munro, A. H. and R. A. J. Wixley. "Estimators Based on Order Statistics of Small Sample from a Three-Parameter Lognormal Distribution," Journal of the American Statistical Association, 65: 212-225 (March 1970).

27. Nagarsenker, B. N. Lecture Material distributed in MA 683, Mathematical Statistics I. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1985.

28. Porter, James E. A Modified Kolmogorov-Smirnov, Anderson-Darling, and Cramer-von Mises Test for the Pareto Distribution with Unknown Scale and Location Parameters. MS Thesis, GSO/ENC/85D. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, Dec 1985.

29. Pritsker, A. Alan B. Introduction to Simulation and SLAM II. New York: Halsted Press, 1984.

30. Rao, C. R. Advanced Statistical Methods in Biometric Research. New York: John Wiley & Sons, Inc., 1952.

31. Ream, Thomas J. A New Goodness of Fit Test for Normality with Mean and Variance Unknown. MS Thesis, GOR/MA/81D-9. School of Engineering, Air Force Instiitiute of Technology (AU), Wright-Patterson AFB OH, Dec 1981.

32. SAS Institute Inc. SAS User's Guide: Statistics, Version 5 Edition. Cary, NC: SAS Institute Inc., 1985

33. Stephens, M. A. "EDF Statistics for Goodness of Fit and Some Comparisons", Journal of the American Statistical Association, 69: 730-737 (Sep 1974).

34. Viviano, Philip J. A Modified Kolmogorov-Smirnov, Anderson-Darling, and Cramer-von Mises Test for the Gamma Distribution with Unknown Location and Scale Parameters. MS Thesis, GOR/MA/82D-4. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, Dec 1982.

35. Woodbury, Larry B. A New Goodness of Fit Test for the Uniform Distribution with Unspecified Parameters. MS Thesis, GOR/MA/82D-6. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, Dec 1982.

36. Woodruff, Brian W. and others. "A Modified Kolmogorov-Smirnov Test for Weibull Distributions with Unknown Location and Scale Parameters," IEEE Transactions Reliability, R-32: 209-213 (Jun 1983).

37. Woodruff, Brian W. and others. "Modified Goodness-of-Fit Tests for Gamma Distributions with Unknown Location and Scale Parameters," IEEE Transactions Reliability, (to appear).

38. Woodruff, Brian W. and others. "Modifed Goodness-of-Fit Tests for the Logistic Distribution with Unknown Location and Scale Parameter," (unpublished).

39. Yoder, John D. Modified Kolmogorov-Smirnov, Anderson-Darling, and Cramer-Von Mises Tests for the Logistic Distribution with Unknown Location and Scale Parameters. MS Thesis, GOR/ENC/83D. School Of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1983.

40. Yuan, Pae-Tsi. "On the Logarithmic Frequency Distribution and Semi-Logarithmic Correlation Surface," Annals of Mathematical Statistics, 4: 30-74 (1933).

VITA

Captain Lynnette Townsend Whitsel was born in Cape Girardeau, Missouri, on 27 July 1960. She graduated from Puxico High School in Puxico, Missouri in 1978 and attended Southeast Missouri State University in Cape Girardeau, Missouri, receiving her degree in Mathematics and Fine Arts in May 1982. Upon graduation, she received a commission in the USAF through ROTC and began active duty in October of that year. From October 1982 to May 1985, Captain Whitsel was assigned to Headquarters Military Airlift Command, Scott AFB, Illinois as an Operation Research Analyst (AFSC C2685) until May of 1985. She then entered the Graduate Operation Research Program, School of Engineering, Air Force Institute of Technology.

Permanent Address: 1716 Themis
Cape Girardeau, Mo. 63701

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
AFIT/GOR/MA/86D-6

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
School of Engineering

6b. OFFICE SYMBOL *(If applicable)*
AFIT/ENS

7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS *(City, State, and ZIP Code)*
Air Force Institute of Technology
Wright Patterson AFB, OH 45433

7b. ADDRESS *(City, State, and ZIP Code)*

8a. NAME OF FUNDING / SPONSORING ORGANIZATION

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS *(City, State, and ZIP Code)*

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*
See Box 19

12. PERSONAL AUTHOR(S)
Lynnette T. Whitsel, B.S., Capt, USAF

13a. TYPE OF REPORT
MS Thesis

13b. TIME COVERED
FROM ________ TO ______

14. DATE OF REPORT *(Year, Month, Day)*
1986 December

15. PAGE COUNT
105

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
12	01	

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Probability Distribution Function, Statistical Test, Goodness-of-Fit, Lognormal Distribution

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Title: A MODIFIED GOODNESS-OF-FIT TEST FOR THE LOGNORMAL DISTRIBUTION WITH UNKNOWN SCALE AND LOCATION PARAMETERS

Thesis Chairman: Dr. Albert H. Moore
Professor of Mathematics

Approved for public release: IAW AFR 190-1.
LYNN E. WOLAVER 5 March 87
Dean for Research and Professional Development
Air Force Institute of Technology (AFC)
Wright-Patterson AFB OH 45433

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Prof. Albert H. Moore

22b. TELEPHONE *(Include Area Code)*
(513) 255-3098

22c. OFFICE SYMBOL
AFIT/ENC

DD Form 1473, JUN 86 — *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

This thesis developed modified goodness-of-fit test for the three parameter lognormal distribution when the location and scale parameters must be estimated from the sample. The criticial values were generated for the Kolmogorov-Smirnov (K-S), the Anderson-Darling (A-D), and the Cramer-von Mises (C-VM) goodness-of-fit tests, using the Monte Carlo methods of 5000 repetitions, to simulate samples of size 5, 10, 15, 20, 25, and 30 and the shape parameter ranged from 1.0 to 4.0 in increments of .5.

The second part of the research also involved a Monte Carlo simulation of 5000 repetitions for sample sizes of 5, 15, and 25. From these observations, the power of the test was determined by counting the number of times the modified goodness-of-fit tests incorrectly accepted the null hypothesis that the distribution was lognormally distributed. The data used in this power comparison came from the lognormal distribution where shape = 1.0 and 3.0, Weibull, gamma, beta, exponential, and normal distributions.

The third and final phase of research was to determine the functional relationship, if any, between the known shape parameter and the new modified critical values. This was completed by using SAS.

DATE
FILME
5-8